The molecular regulation of terpene polymorphism in *Thymus vulgaris*

DISSERTATION

zur Erlangung des akademischen Grades

Doktor rerum naturalium

(Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät

der Friedrich-Schiller-Universität Jena

von Julia Nina Christina Asbach, Dipl. Biol.

geboren am 03. Mai 1977 in Hamburg

Gutachter:

1. Prof. Dr. Jonathan Gershenzon
   Max Planck Institut für Chemische Ökologie, Jena, Deutschland

2. Prof. Dr. Severin Sasso
   Friedrich-Schiller-Universität Jena, Deutschland

3. Prof. Dr. Bodil Ehlers
   Universität Aarhus, Dänemark

Tag der öffentlichen Disputation: 17. November 2014

TABLE OF CONTENTS

1 General Introduction 1

Terpene biosynthesis in plants 2

Thymus vulgaris – The system 8

Questions addressed in this thesis 14

2 Terpene composition of the *Thymus vulgaris* chemotypes investigated in this thesis 15

Introduction 15

Material and Methods 20

Results 22

Discussion 27

3 Characterization of terpene biosynthetic genes of *Thymus vulgaris* 31

Introduction 31

Material and Methods 33

Results 40

Discussion 65

4 Molecular control of the monoterpene polymorphism in *Thymus vulgaris* 83

Introduction 83

Material and Methods 86

Results 89

Discussion 99

5 Plant regeneration from *Thymus vulgaris* leaf explants 107

Introduction 107

Material and Methods 109

Results 111

Discussion 117

6 General Discussion 131

7 Summary 141

8 Zusammenfassung 143

9 References 147

10 Acknowledgments 169

11 Supplementary Material 173

12 Eigenständigkeitserklärung 181

13 Curriculum Vitae 183

1 General Introduction

Plant secondary metabolites constitute a large group of structurally diverse compounds that are also referred to as specialized metabolites or natural products. Traditionally secondary metabolites are defined as compounds that, unlike primary metabolites, are not required for basic survival but which none the less provide a competitive advantage to the organism (Theis and Lerdau 2003). Biochemically, secondary metabolites can be divided into three major groups: terpenoids, phenolic compounds, and nitrogen-containing compounds, of which the latter can be further subdivided into glucosinolates, cyanogenic glucosides, and alkaloids. Of the more than 200.000 plant secondary metabolites, many have been found to be of ecological importance to plants, serving as attractants for pollinators or seed dispersers, defenses against herbivores and pathogens, or as allelochemicals against competitors. The functional diversity of secondary metabolites may be demonstrated best by terpenes.

Terpenoids, also called isoprenoids or terpenes, constitute the largest and most diverse class of secondary metabolites, and conservative estimates suggested that at least 65.000 different terpenoids exist in nature, of which 43.000 have already been described (Bohlmann and Keeling 2008, Oldfield and Lin 2012, Hamberger and Bak 2013). While the majority has been isolated from plants, they appear to be almost ubiquitous in nature, occurring also in bacteria, fungi, marine organisms, mollusks or even fish, insects, and mammals (Gershenzon and Dudareva 2007). Given their wide distribution and diversity and the fact that most have no apparent function in the basic processes of growth and development, it is not surprising that they have been historically referred to as secondary metabolites. Their biological roles, however, are as manifold as their structures are diverse, and include physiological, structural and ecological functions. A few specialized groups of terpenes with well-characterized functions include (i) the carotenoids, such as xanthophylls, carotenes, lycopene, *all-trans*-retinal, and retinol (a vitamin A), which play important roles in photosynthetic light-harvesting, photoprotection, photoreception, cell proliferation and differentiation, as well as visual attraction, (ii) the steroids, such as cholesterol, an important component of cell membranes and the different steroid hormones, like testosterone, estrogens, progesterones, cortisol, and vitamin D and analogs (e.g. calcitriol), (iii) ubiquinone, a proton- and electron carrier of the respiratory chain, (iv) phytol, the side chain

of the photosynthetic light-harvesting pigment chlorophyll, (v) diverse plant hormones such as abscisic acid, gibberellins, and strigolactones which regulate plant growth and influence various developmental processes, and other vitamins, such as (vi) *α*-tocopherol (a vitamin E) (Figure 1-1). Furthermore there are several proteins, such as heme A or G proteins, which are bound to the cell membrane by isoprenoid anchors. Besides their preeminent ecological and physiological roles, terpenoids have also been of interest to humankind since antiquity as flavors and fragrances as well as pharmaceuticals.

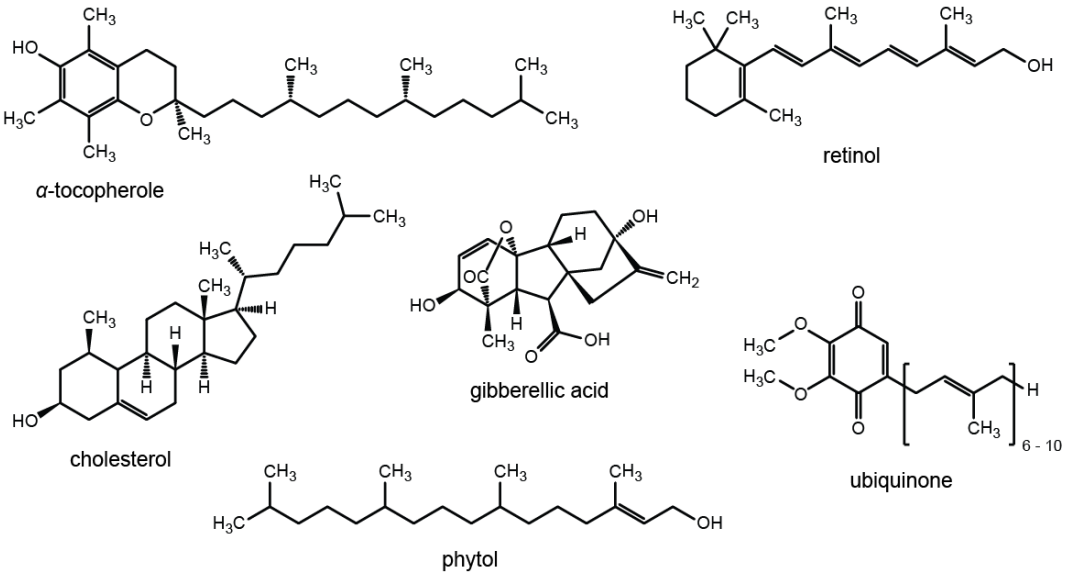


Figure 1-1: A selection of terpenoids with well-known physiological and biological activities.

Terpene biosynthesis in plants

The structural diversity exhibited by terpenoids and the related wide range of physiological activities might lead one to believe that these secondary metabolites originate from complex precursors. However, terpenoids are derived from two universal five-carbon precursors, isopentenyl diphosphate (IPP) and its allylic isomer, dimethylallyl diphosphate (DMAPP). In plants, two independent and spatially separated pathways – the methylerythritol phosphate (MEP) pathway and the mevalonic acid (MVA) pathway – are responsible for the formation of these C_5-isoprene building blocks (Figure 1-2).

The C_5 units of the MEP pathway give rise principally to hemiterpenes (C_5), monoterpenes (C_{10}), diterpenes (C_{20}), and carotenoids (C_{40}), while the C_5 units of the MVA pathway lead to sesquiterpenes (C_{15}), triterpenes (C_{30}), and sterols. The MEP pathway is

considered to be exclusively plastidic, while the MVA pathway appears to be distributed between the cytosol, the endoplasmic reticulum, and peroxisomes (Dudareva et al. 2013). The MEP pathway involves seven enzymatic steps and uses, in contrast to the MVA pathway, both D-glyceraldehyde-3-phosphate (GAP) and pyruvate (Pyr) as precursors to form IPP and DMAPP. The MVA pathway on the other hand consists of six enzymatic reactions, and is initiated by a stepwise condensation of three molecules of acetyl-CoA, followed by a reduction, two subsequent phosphorylations, and a decarboxylation step to yield IPP as the final product. IPP can be converted to and from DMAPP by isopentenyl diphosphate isomerase.

Both IPP and DMAPP are utilized as substrates by prenyltransferases to produce the prenyl diphosphates, geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP), used in turn by the large family of terpene synthases. GPP, the C_{10} precursor of monoterpenes, and GGPP, the C_{20} precursor of diterpenes, are formed via the head-to-tail condensation of one DMAPP molecule with one or three IPP molecules, respectively. Likewise FPP, the C_{15} precursor of sesquiterpenes, is formed by the head-to-tail condensation of one DMAPP molecule with two IPP molecules.

In plants the isoprenoid pathway network is compartmentalized in different subcellular compartments, such as the cytosol, endoplasmic reticulum, plastids, and the mitochondria (Figure 1-3). IPP produced in plastids via the MEP pathway is considered the precursor of isoprenoid compounds produced in the plastids, such as hemi-, mono- and diterpenes, as well as carotenoids (tetraterpenes), and chlorophylls, while isoprenoid compounds synthesized in the cytosol and mitochondria, such as sesquiterpenes, triterpenes, and sterols incorporate IPP produced via the cytosol-specific MVA pathway (Nagegowda 2010, Vranová et al. 2012). This subcellular localization, however, is not absolute and a certain amount of metabolic crosstalk is known to occur primarily from the plastids to the cytosol. Thus IPP and other short prenyl diphosphates, such as DMAPP, GPP, and FPP, can be translocated through the plastid membrane, while higher prenyl diphosphates, such as GGPP are not transported with a comparable efficiency. This exchange of metabolites between cellular compartments is known to be a highly regulated process. Besides the transport of the short prenyl diphosphate precursors between different compartments, the localization of terpenoid biosynthesis is also influenced by the intracellular localization of the different terpenoid synthases. Most mono- and diterpene synthases are localized via a short N-terminal sequence to the plastids, while most sesquiterpene synthases lack this transit

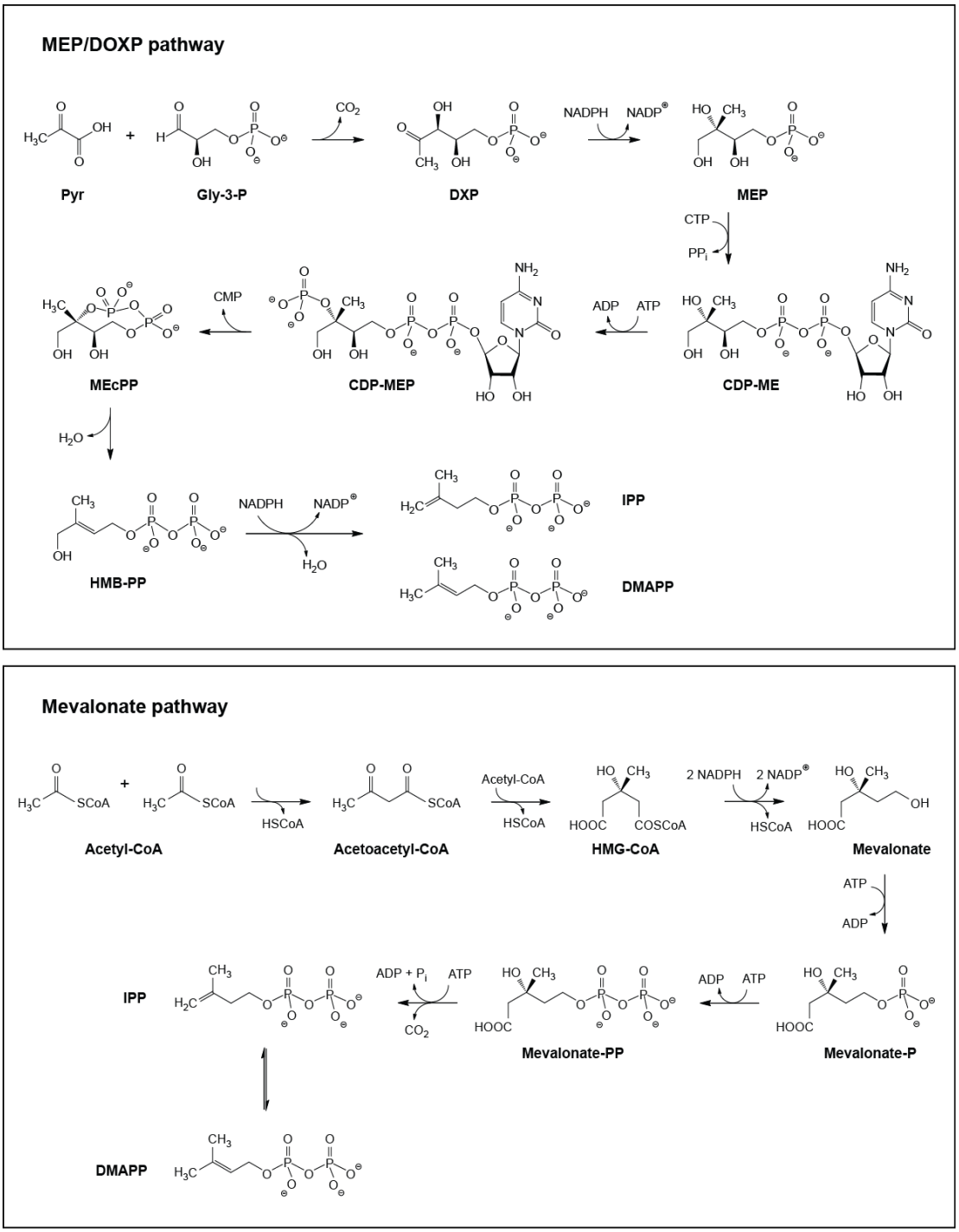


Figure 1-2: Biosynthesis of the C5 intermediates IPP and DMAPP. Compounds are: Acetyl-Coenzyme A (**Acetyl CoA**), Acetoacetyl-Coenzyme A (**Acetoacetyl-CoA**), 4-(cytidine 5'-diphospho)-2*C*-methyl-D-erythritol (**CDP-ME**), 4-(cytidine 5'-diphospho)-2*C*-methyl-D-erythritol-2-phosphate (**CDP-MEP**), dimethylallyldiphosphate (**DMAPP**), 1-deoxy-D-xylulose-5-phosphate (**DXP**), D-glyceraldehyde-3-phosphate (**Gly-3-P**), 1-hydroxy-2-methyl-2-(*E*)-butenyl-4-phosphate (**HMB-PP**), 3-

hydroxy-3-methylglutaryl-CoA (**HMG-CoA**), isopentenyl diphosphate (**IPP**), 2C-methyl-D-erythritol-2,4-cyclodiphosphate (**MEcPP**), 2*C*-methyl-D-erythritol-4-phosphate (**MEP**), mevalonic acid (**Mevalonate**), mevalonate-5-phosphate (**Mevalonate-P**), mevalonate-5-diphosphate (**Mevalonate-PP**), pyruvate (**Pyr**).

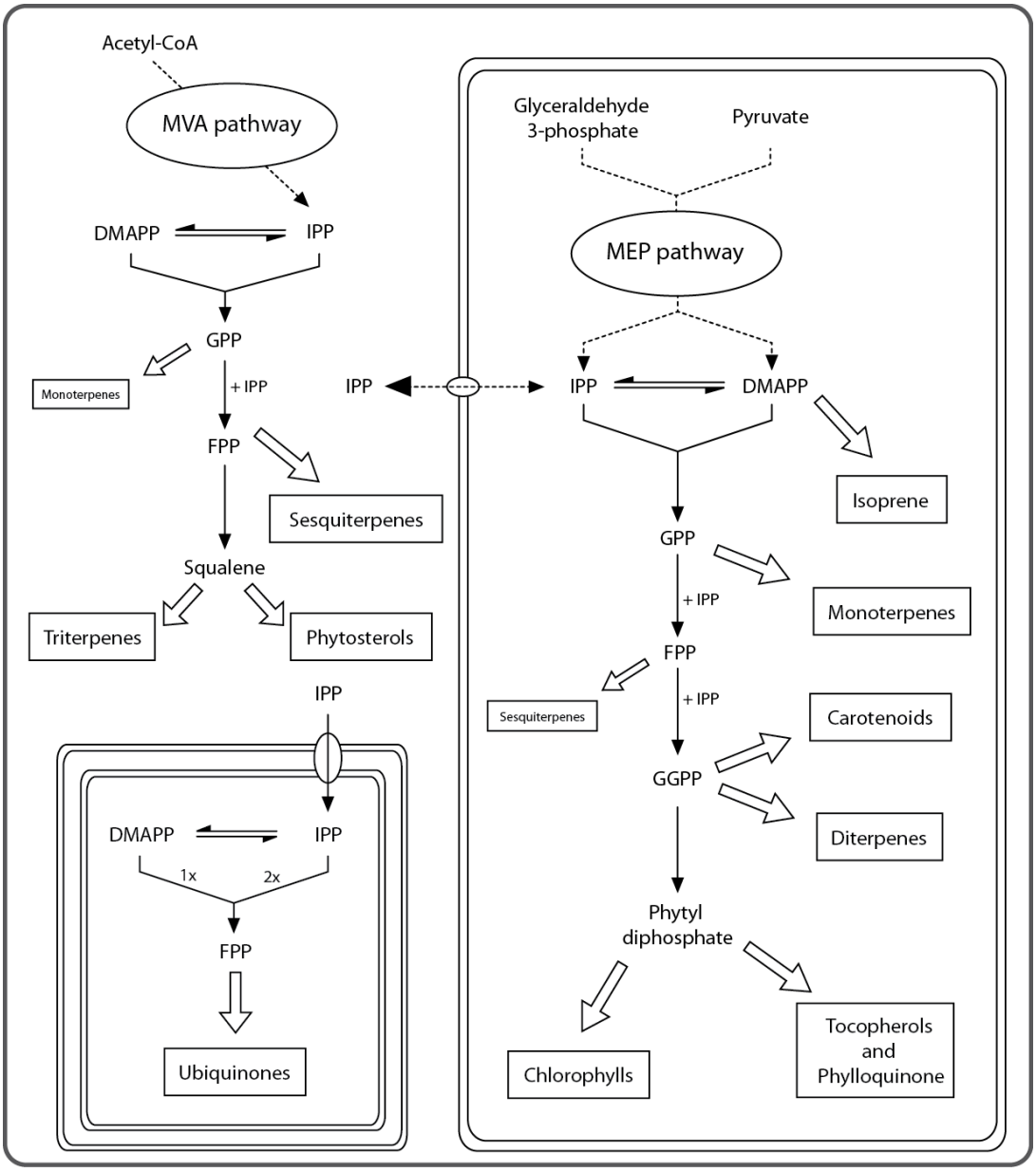


Figure 1-3: Subcellular compartmentation of plant terpenoid biosynthesis. The mevalonic acid pathway is located in the cytosol, in peroxisomes, and in the endoplasmic reticulum. The methyl-erythritol-phosphate pathway is located in the plastids. Abbreviations are as follows: MVA: mevalonic acid, MEP: 2-C-methyl-D-erythritol 4-phosphate, DMAPP: dimethylallyl pyrophosphate, IPP:

isopentenyl pyrophosphate, GPP: geranyl pyrophosphate, FPP: farnesyl pyrophosphate, GGPP: geranylgeranyl pyrophosphate.

peptide and are localized to the cytosol. However, some terpene synthases have been described that deviate from this pattern, including a bifunctional nerolidol/linalool synthase from strawberry (*Fragaria* x *ananassa*), which lacks a plastid-targeting sequence and is localized to the cytosol and a santalene and bergamotene-producing sesquiterpene synthase from the wild tomato *Solanum habrochaites* which possesses a chloroplast targeting sequence and is localized in the plastids (Sallaud et al. 2009, Nagegowda 2010, Vranová et al. 2012).

Terpene synthases

The large number of terpene synthases (TPS) and the ability of these enzymes to form multiple products from a single prenyl diphosphate substrate are the main reasons for the Brobdingnagian diversity of terpenoids found in plants (Degenhardt et al. 2009). Most TPS use prenyl diphosphate substrates with *trans*-configurations, such as GPP, FPP, and GGPP though some recently discovered enzymes use substrates which have a *cis*-configuration at one of the double bonds (Dudareva et al. 2013).

The TPS gene family has been divided into seven subfamilies (designated TPS-a, TPS-b, TPS-c, TPS-d, TPS-e/f, TPS-g, and TPS-h) based on phylogenetic sequence analysis, functional assessment, and gene architecture (Chen et al. 2011). The enzymatic reaction mechanism of TPS is initiated by the formation of a highly reactive carbocation, which is realized either via an ionization-dependent strategy or a proton-dependent strategy. The ionization-dependent strategy, similar to the approach used by the farnesyl diphosphate synthase, is employed by so called class I TPS, and initiates the enzymatic reaction by metal-activated ionization of the prenyl diphosphate to form highly reactive carbocation intermediates. These ionization-dependent terpenoid cyclases – monoterpene, sesquiterpene and diterpene cyclases – contain the DDxxD/E and NSE/DTE signature motifs in their α-domains, which together bind the trinuclear magnesium cluster that triggers the departure of the substrate diphosphate leaving group (Christianson 2006). The proton-dependent strategy, employed by certain diterpene cyclases, forms copalyl diphosphate (CDP) as either the final product or an intermediate. These class II TPS contain a DxDD signature motif in their $\gamma\beta$-domain, in which the central aspartate residue is utilized as the proton donor that triggers initial carbocation formation (Christianson 2006, Chen et al. 2011). Bifunctional class I/II enzymes, such as the diterpene cyclase abietadiene synthase from grand fir (*Abies grandis*),

contain both class I and class II functional domains and bifunctional properties (Christianson 2006, Chen et al. 2011).

Terpenoid diversity is further increased by enzymatic modifications, such as oxidation, peroxidation, hydroxylation, dehydrogenation, methylation, acylation, cleavage, or glycosylation, which may also change or alter their biological activities (Chen et al. 2011, Dudareva et al. 2013). A class of enzymes playing an important role in these subsequent pathway steps is the widespread family of cytochrome P450 monooxygenases.

Monooxygenases

Similar to the terpenoid synthases, cytochromes P450s (P450) form one of the oldest and largest class of enzymes and can be found in all biological kingdoms, from archaeae, protists, fungi, plants and animals to even viruses (Nelson and Werck-Reichhart 2011). Although the reaction most often catalyzed is a hydroxylation (insertion of oxygen), their reaction spectrum is manifold and dehydrogenation, reduction, isomerization, dimerization, carbon-carbon bond cleavage, as well as N-, O-, and S-dealkylations, sulphoxidations, deaminations, and desulphurations have been described to date (Weitzel and Simonsen 2013). The P450s found in land plants are currently assigned to 127 families in 11 phylogenetically distinct clans. These clans can be divided into two groups: single-family clans (CYP51, CYP74, CYP97, CYP710, CYP711, CYP727, and CYP746) and multi-family clans (CYP71, CYP72, CYP85, and CYP86). Single family clans usually code for enzymes involved in essential functions, such as sterol formation for membranes (CYP51) and phytohormones for plant development and plant defense (CYP710 and CYP74), and appear to be under purifying selection as well as being constrained from duplication. Multi-family clans on the other hand, such as CYP71 which comprises 54 families and over half of the P450s described in higher plants, are less restricted and encompass a variety of functions, ranging from essential functions as the biosynthesis of plant hormones and biopolymers such as lignin and cutin, to the specialized metabolism of aromatic and aliphatic amino acid derivatives (e.g. indolic derivatives, phenylpropanoids, glucosinolates, and cyanogenic glucosides), alkaloids, and terpenoids and their derivatives (Nelson and Werck-Reichhart 2011, Hamberger and Bak 2013, Weitzel and Simonsen 2013).

Terpenoid-oxidizing P450s were among the first P450s identified in plants and belong to the oldest existing plant P450 families. Most of these classical oxygen-requiring plant P450s are dependent on the presence of cofactors, such as the nicotinamide adenine dinucleotides NADH and NADPH, which are essential for the electron transport. These

cofactors can be either covalently bound to the active site of the enzyme, forming a prosthetic group, or act as loosely bound co-substrates. The plant monooxygenases, which catalyze the hydroxylation of monoterpenes, are usually moored, via their N-terminal signal anchor sequence buried in the lipid bilayer, to the endoplasmic reticulum and with their catalytic domain in the cytosol. These in turn are equipped with a FAD flavoprotein reductase component to catalyze the electron release from NAD(P)H and an iron-sulphur protein which acts as an electron carrier between the flavoprotein and the cytochrome P450 component (Crocoll 2011, Schuler and Rupasinghe 2011).

Thymus vulgaris – The system

Thymus vulgaris L. belongs to the Labiatae family (Lamiaceae), subfamily Nepetoideae, tribe Menthae and is a native of southern Europe. It is an aromatic, perennial dwarf subshrub, 10 – 30 cm in height with slender, wiry and spreading branches. The small linear, gland-dotted, evergreen leaves with curved leaf margins, and the two-lipped white to pale violet flowers in axillary clusters on the branchlets or terminal oval or rounded heads, are well known (Cullen et al. 2000).

Several explanations exist concerning the origin of the name 'thymus'. It is assumed that the Latin name *Thymus* either derives from the Greek words [θύω] *thyo* and [θύειν] *thyein* respectively (to perfume, burn incense, to offer an incense sacrifice) or [θυμός] *thymos* (strength, courage). Originally 'thymus' described a group of aromatic plants that were used as stimulants of vital functions and belonged either to the genera *Thymus* or *Satureja*.

The widespread use of thyme dates back to ancient Egypt where various species were grown to perfume unguents (soothing ointments spread on skin injuries) and for embalming, and possibly for medical purposes as well. Evidence from the first century AD indicates that the Greeks and Romans used it in the same way. In his work about medicinal plants and poisons (Περί ὕλης ἰατρικής – lat. *De Materia Medica*, translation of Berendes, 1902) Dioscorides (first century AD) writes about [ἔρπύλλος] *erpyllos* and [θυμός] *thymos*, while Pliny the Elder (first century AD) mentions 'thymium' in his work *Naturalis historiae*. The known pharmacological effect of thyme include: antibacterial, antifungal, antiviral, and spasmolytical effects (Stahl-Biskup and Sáez 2002). The antibacterial effects inhibit a broad spectrum of bacteria, with gram-positive bacteria like *Bacillus cereus* or *Staphylococcus aureus*, being more sensitive than gram-negative bacteria like *Escherichia coli*, *Helicobacter*

pylori or *Pseudomonas aeruginosa*. Antifungal effects could be demonstrated on several food-spoiling yeasts, such as *Saccharomyces cerevisiae*, *Geotrichum candidum*, and *Brettanomyces anomalus* (Conner and Beuchat 1984, Kunicka-Styczyńska 2011), as well as on several dermatophytes and some phytopathogenic fungi, such as *Colletotrichum lindemuthianum*, *Fusarium solani*, *Pythium ultimum*, and *Rhizoctonia solani* (Stahl-Biskup and Sáez 2002, Tullio et al. 2007). The essential oil of *T. vulgaris* has furthermore been successfully used in agricultural seed preservation as it inhibits aflatoxin production and mycelial growth of *Aspergillus flavus* or *A. parasiticus* without producing significant phytotoxic effects on the germination and growth of the treated seeds. These pharmacological effects of thyme can all be attributed to the essential oil, which is rich in monoterpenoids.

The mode of action of thyme monoterpenes on microorganisms is thought to arise from their ability to pass through cell walls and cytoplasmic membranes, disrupt the structure of their different layers of polysaccharides, fatty acids and phospholipids, and permeabilize them (Bakkali et al. 2008). In bacteria, this permeabilization of the membranes is associated with the loss of ions and the reduction of membrane potential, the collapse of the proton pump, and the depletion of the adenosine triphosphate (ATP) pool, resulting in cell death (Helander et al. 1998, Bakkali et al. 2008). The two thyme phenolic monoterpenoids thymol and carvacrol, for example, have been shown to bind to amine groups of the bacterial membrane proteins, leading to an increased membrane fluidity which results in an efflux of ions and ATP, as well as a disturbed membrane potential and pH gradient (Stahl-Biskup and Sáez 2002, Bakkali et al. 2008, Radulovic et al. 2013). Similar mechanisms appear to be the cause of fungal growth inhibition and fungal hyphae degeneration (Stahl-Biskup and Sáez 2002, Bakkali et al. 2008). Depending on the monoterpene composition, the essential oil of the various *T. vulgaris* chemotypes, which are described in more detail below, differ in their antimicrobial activity. Thus, the most active oil is of the thymol chemotype followed by the oils of the carvacrol, geraniol, and linalool chemotypes, while the *α*-terpineol and thujanol-4 chemotypes are the least active (Stahl-Biskup and Sáez 2002). The antimicrobial activity of monoterpenes can be attributed to their functional groups, where alcohols appear to have the highest activity, followed by aldehydes, and then ketones. The high activity of thymol and carvacrol against many microorganisms can be ascribed to the phenolic hydroxyl group as well as to the phenolic ring itself (thymol is more active than *p*-cymene), while the relative position of the hydroxyl group also influences the activity (thymol is more active than carvacrol). The presence of an ester group in the monoterpenoid structure also increases the

activity in comparison to the activity of its parent compound; thus geranyl acetate is more active than geraniol (Bakkali et al. 2008, Saad et al. 2013). However, the biological activity of essential oils as a whole is often higher than the sum of their individual components. This synergism has been demonstrated for several components, such as carvacrol or thymol (Radulovic et al. 2013, Saad et al. 2013).

Trichomes

In thyme the terpene-rich essential oil is produced and stored in so called glandular trichomes which are situated on the aerial parts of the plant. These types of specialized secretory cells appear to be an ancient invention of vascular plants, as fossilized leaf materials from seed ferns of the late Paleozoic and the late Carboniferous would suggest, and have arisen in plants independently several times, which make them common in some genera while being absent or rare in others (Tissier 2012, Lange and Turner 2013). The metabolic functions of the specialized cells of these epidermal hairs are manifold and include the biosynthesis and secretion of nectar, mucilage, digestive enzymes, and protective secondary metabolites, while others have a role in the absorption of nutrients or the secretion of salts from the plant tissue (Lange and Turner 2013).

The glandular trichomes of the Lamiaceae are well adapted to retain and store volatile compounds, such as mono- and sesquiterpenes, which would otherwise simply diffuse into the headspace. In thyme, the major structural type of glandular trichomes that store essential oil, the peltate glandular trichomes, arise as epidermal protuberances that divide asymmetrically to produce a basal cell (BC), a stalk cell (SC), and an apical disk initial cell. An additional series of up to seven anticlinal divisions of the apical cell lead to 10 – 14 secretory disk cells (DC), which form the head of the peltate glandular trichomes (Figure 1-4). To prevent backflow of the secreted substances back through the apoplast, the cell wall of the SC appears to be enveloped by a Casparian-like strip on the lateral and a portion of the basal part of the cell. The terpenoid containing essential oils are produced in the DCs and secreted into the sub-cuticular space, where they are stored. The cuticle ruptures when, for instance an insect treads upon the trichome or when the leaves are crushed, and the stored volatile compounds are then released (Bruni and Modenesi 1983, Yamaura et al. 1992).

Research has verified a substantial metabolic specialization of the glandular cells of trichomes from several plant families. This combination of specificity and high-level expression of trichome pathways turns these cells into treasure troves for the discovery,

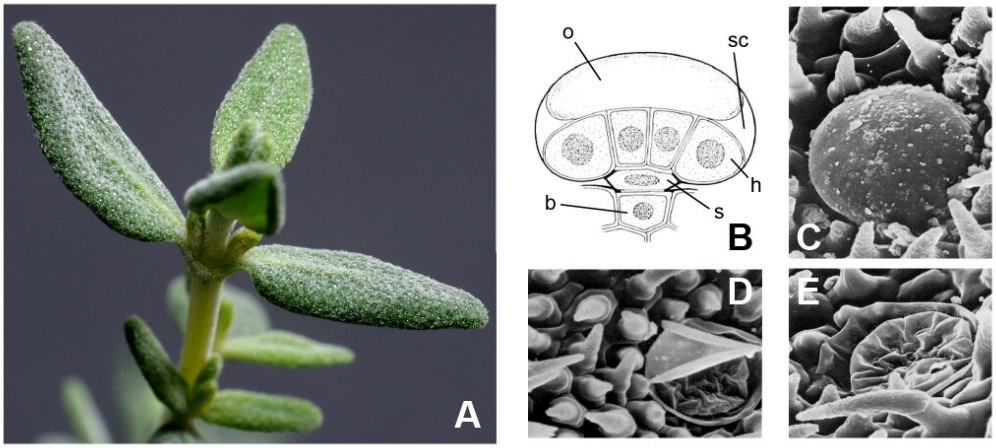


Figure 1-4: Peltate glandular trichomes of *Thymus vulgaris*. (**A**) Photograph of *T. vulgaris* leaves showing the glandular trichomes as reflecting dots on the foliar surface. (**B**) Schematic drawing of a peltate glandular trichome of Lamiaceae. (b) basal cell, (h) head cells (s) stalk cells (sc) subcuticular cavity, (o) oil drop. (Modified after Fahn (1988)). (**C – E**) SEM micrographs showing mature *T. vulgaris* glandular trichomes.

isolation, and characterization of genes involved in those pathways (Tissier 2012, Lange and Turner 2013). In the case of the Lamiaceae, the isolation of these specialized DCs has helped to identify genes directly involved in the terpenoid pathway, of which the elucidation of the *p*-menthane monoterpene pathway in *Mentha* sp., including three monooxygenases, is probably the best known example (Lange and Turner 2013). However even though mint is the genus whose essential oil biosynthesis has been most extensively studied, there are numerous other examples within the Lamiaceae, including six monoterpene synthases and four sesquiterpene synthases from sweet basil (*Ocimum basilicum*), four different monoterpene synthases as well as a putative monooxygenase from different *Perilla* sp., three monoterpene synthases, three sesquiterpene synthases, and four monooxygenases from oregano (*Origanum vulgare*)(Crocoll 2011), as well as several mono- and sesquiterpene synthases from lavender (*Lavandula* sp.), sage (*Salvia* sp.) and others (Lange and Turner 2013).

Polymorphisms in thyme

Two different polymorphisms have been described for thyme. **The first polymorphism** is a gender dimorphism concerning the breeding system of the plant. In natural populations self-compatible hermaphrodites co-exist with male-sterile individuals (hereafter ‘female’), which bear smaller flowers that either lack anthers or have rudimentary anthers that do not bear pollen. The genetic basis of this gynodioecy in thyme involves both

cytoplasmic and nuclear control, involving two sets of genes (Couvet et al. 1986, Tarayre et al. 1997, Delph et al. 2007). One set is located in the mitochondrial genome and prevents plants from producing pollen (i.e. turning them into females), while the other set is in the nucleus and restores male fertility, thus turning plants back into hermaphrodites (Delph et al. 2007). In gynodioecious species the females are obligate outcrossers while the hermaphrodites may self which prevents the deleterious effects of excessive inbreeding. This polymorphism is stable and thought to be influenced by strong ecological constraints (Thompson and Tarayre 2000, Ehlers and Thompson 2004b). **The second polymorphism** in thyme is chemical affecting the essential oil composition of individual plants. In the south of France, six different chemical forms of thyme were originally described that differ in the monoterpene composition of the essential oil. These so-called chemotypes, geraniol (G), *α*-terpineol (A), *trans*-sabinene hydrate or thujanol-4 (U), linalool (L), carvacrol (C), and thymol (T), are named after the dominant monoterpene found in the essential oil (Passet 1971, Vernet et al. 1986). An investigation of the chemical polymorphism in Spanish populations revealed a seventh chemotype, 1,8-cineole (E), while the geraniol chemotype could not be detected among the investigated plants (Adzet et al. 1977). More recently this seventh chemotype was also found in French populations (Keefover-Ring et al. 2009). The seven chemotypes can be grouped into phenolic (C and T) and non-phenolic (G, A, U, L, and E) chemotypes depending on the molecular structures of their dominant monoterpenes. The genetic control of this chemical polymorphism has been analyzed by controlled crosses and sampling from natural populations (Vernet et al. 1986). It appears that the variation is determined by an epistatic series of five loci giving a set order of chemotype dominance (G → A → U → L → C → T). Two loci code for the G chemotype, while the remaining chemotypes are coded by a single pair of alleles at each locus. This means that a plant with the dominant G allele will have the G phenotype, regardless of whether it possesses dominant or recessive alleles at the remaining loci. If a plant is homozygous recessive at the G and A loci, but has a dominant U allele, then it will have the U phenotype, and so on down the chain. A plant homozygous recessive at all five loci has a T phenotype. In Spanish populations the E chemotype appears to take the place in genetic dominance of that of the G chemotype, which itself is absent in these populations (Adzet et al. 1977, Vernet et al. 1986). The overall composition of the essential oil of an individual plant, however, has been reported to be unaffected by seasonality or changes in environmental conditions (Vernet et al. 1986).

The chemical variation may represent adaptions to different environmental factors since the phenolic chemotypes dominate thyme populations in hot dry sites close to the Mediterranean sea, while the non-phenolic chemotypes are found further inland, particularly above 400 m elevation, i.e. in wetter, cooler climates (Passet 1971, Granger and Passet 1973). The non-phenolic chemotypes are also more adapted to low temperatures in winter while phenolic chemotypes are sensitive to extremely cold temperatures. Phenolic chemotypes however, appear to be better adapted to summer drought than non-phenolic chemotypes (Amiot et al. 2005, Thompson et al. 2007, Thompson et al. 2013). The effects of chemical variation in *T. vulgaris* on the environment, such as allelophatic influence on co-occurring plants and soil organisms, or deterrence on herbivores, have also been investigated on several occasions (Linhart and Thompson 1999, Ehlers and Thompson 2004a, Linhart et al. 2005, Ehlers et al. 2012).

Despite the fact that chemical polymorphism in thyme has been a subject of study for over thirty years and numerous insights into its heredity and ecological role could be gleaned, many questions still remain unanswered.

Questions addressed in this thesis

To understand more about the evolutionary origin and ecological significance of monoterpene polymorphism in *Thymus vulgaris*, some key enzymes of monoterpene biosynthesis were isolated, characterized, and their role in the chemotype polymorphism investigated. Chapter I covers the chemical and genetic description of the chemotypes used in this thesis. Chapter II focuses on the molecular basis of the polymorphism and describes the isolation and identification of several enzymes involved in the production of major essential oil components, as well as their possible phylogenetic relationship. In Chapter III the information obtained in the two preceding chapters was used to investigate the genetic control underlying the chemical polymorphism. Chapter IV delineates the establishment of a regeneration protocol for *T. vulgaris* as a basis for the development of a plant specific transformation protocol. Transformation of thyme with the monoterpene biosynthetic genes isolated in this thesis could help to prove some of the genetic control mechanisms identified.

2 Terpene composition of the *Thymus vulgaris* chemotypes investigated in this thesis

Introduction

Plant secondary metabolites are characterized not only by their great diversity of chemical structures, but also by their variable distribution in the plant kingdom. Each species or group of related species often has its own complement of secondary metabolites. For example, nitrogen containing tropane alkaloids are widely distributed in the plant family of the Solanaceae including angel's trumpet (*Datura* sp.), deadly nightshade (*Atropa belladonna*), but also potato (*Solanum tuberosum*), tomato (*Solanum lycopersicum*) or tobacco (*Nicotiana tabaccum*), but are rare in most other plant groups. The glucosinolates are mostly restricted to members of the Brassicales such as *Arabidopsis thaliana*, *Brassica oleracea* (e.g. cabbage, broccoli, and cauliflower), or mustard (*Sinapis* sp.) (Wink 2010). Sometimes there is even variation among or within populations of a single species. Scientists have always wondered what causes this variation and the availability of modern molecular methods make new approaches to this problem possible.

One of the major groups of secondary metabolites is the terpenes, which includes substances ranging from low to high molecular weight. The low molecular weight terpenes, the C_{10} (monoterpenes), C_{15} (sesquiterpenes) and C_{20} (diterpenes), are commonly stored as oils or resins in the plant. Several plant families, such as the Pinaceae (Keeling and Bohlmann 2006), the Rutaceae (Lota et al. 2001), the Myrtaceae (Stefanello et al. 2011), the Asteraceae (Chadwick et al. 2013), and the Lamiaceae (Lange and Turner 2013) are well known for their terpenoid containing oils and resins.

The Lamiaceae plant family comprises abundant species of economical value for humans, such as species of mint, basil, sage, oregano, and thyme, and is well recognized for the diversity of the secondary metabolites that are synthesized and stored in specialized compartments called glandular trichomes which are found on the aerial parts of the plant (Iijima et al. 2004b). Glandular trichomes are widely distributed in the plant kingdom, and can be single-celled or multicellular, and constitute a wide range of shapes and structures. Their common denominator is the presence of metabolically active cells which can secrete or

store large quantities of specialized metabolites. The compounds produced by glandular trichomes can be hydrophilic, lipophilic, proteins, and poly- or monosaccharides. This type of trichome occurs most frequently in angiosperms, and several dicotyledonous plant families such as the Lamiaceae, Solanaceae, or Asteraceae, are particularly rich in them (Tissier 2012). The peltate glandular trichomes of the Lamiaceae are well known for the production of their terpenoid rich essential oils which are of great economic value. In *Thymus vulgaris* these epidermal oil glands are composed of 10 – 14 radially distributed secretory cells atop a single stalk cell and a basal cell embedded in the plant surface (Figure 1-4). The secretory cells themselves are surmounted by a sub-cuticular cavity into which the oil is secreted and ultimately stored (Bruni and Modenesi 1983).

The terpenes of Lamiaceae species are known to show prominent intraspecific variation. Nearly 60 years ago, chemotypes (or chemical forms) were described by Penfold and Willis (1953) as 'plants in naturally occurring populations which cannot be separated on morphological evidence, but which are readily distinguished by marked differences in the chemical composition of their essential oils'. Intraspecific terpene variation has been reported, among others, for *T. vulgaris* (Passet 1971), *Rosmarinus officinalis* (Granger and Passet 1973), *Mentha spicata* (Kokkini and Vokou 1989), *Origanum vulgare* (Vokou et al. 1993), *Perilla frutescens* (Yuba et al. 1995), and *T. praecox* subsp. *arcticus* (Schmidt et al. 2004). The presence of intraspecific chemotype variation appears to be especially common in the genus *Thymus* (Stahl-Biskup and Sáez 2002, Iijima et al. 2004b). However, terpene variation is not restricted to the Lamiaceae but has also been described for several other plant families, for example the Myrtaceae, with *Angophora* (Leach and Whiffin 1989), *Eucalyptus* (Simmons and Parsons 1987), and *Melaleuca* (Keszei et al. 2008) among others, as well as the Pinaceae (Katoh and Croteau 1998). Since plants that differ in their essential oil composition are often found growing side by side, genetic factors seem to be responsible (Trilles et al. 2006). However, in the cases of *T. vulgaris* and *T. praecox* subsp. *arcticus* a geographic trend in the distribution of chemical variation has been observed (Thompson et al. 2003, Schmidt et al. 2004, Amiot et al. 2005). Although terpene polymorphism in plants has been known for a considerable time, there is almost no information about the precise molecular mechanisms responsible, either in the Lamiaceae or the rest of the plant kingdom.

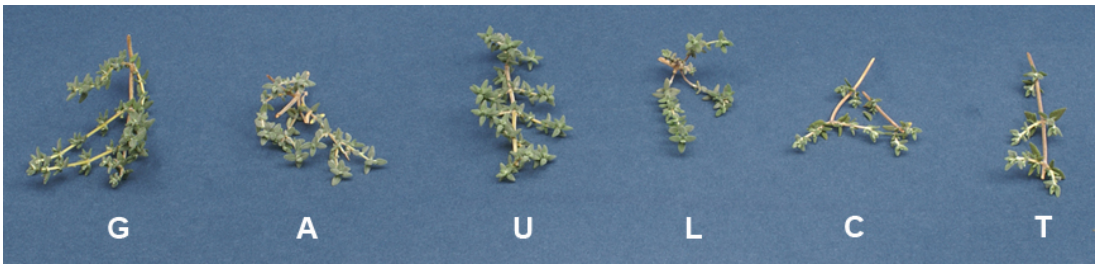


Figure 2-1: **Photograph of twigs of *T. vulgaris* chemotypes used in this thesis and grown in the greenhouse.** Abbreviations are as follows: **G**: geraniol chemotype G_{10}, **A**: α-terpineol chemotype A_{10}, **U**: thujanol chemotype U_6, **L**: linalool chemotype L_{48}, **C**: carvacrol chemotype C_{10}, **T**: carvacrol/thymol chemotype T_{28}.

In the Lamiaceae, intraspecific variation in terpene composition has been best studied in *T. vulgaris*, also known as 'wild thyme' or 'common garden thyme'. Since the early 1960's two polymorphisms occurring in this plant species have been subject to several studies performed in Montpellier, France (Thompson et al. 1998). The first polymorphism is gynodioecy, which involves the coexistence in natural populations of plants bearing either female or hermaphrodite flowers, but not both. The second polymorphism involves the variation in the monoterpene composition and is the focus of this thesis. Seven distinct chemotypes of thyme have been described to date. The essential oil of thyme is in general a complex mixture of about 20 mono- and sesquiterpenes with monoterpenes comprising the majority of the oil, while the sesquiterpenes play only a minor role (Thompson et al. 2003). Thus the aforementioned chemotypes are defined and named based on the dominant monoterpene(s) found in the essential oils of the plants. The geraniol (G), the linalool (L), and the α-terpineol chemotype (A) contain 90 – 95 % of the denominating monoterpene in the essential oil, while the thujanol chemotype (U) contains up to 60 – 65 % of a combination of 4-thujanol, terpinen-4-ol, linalool, and 8-myrcenol. In the carvacrol chemotype (C) on the other hand the essential oil is made up of up to 85 % carvacrol, while the thymol found in the thymol chemotype (T) amounts to no more than 65 % (Granger and Passet 1971, Thompson et al. 2003). The seventh chemotype 1,8-cineole (E), which was first identified in Spain by Adzet et al. (1977) and later also identified in southern France (Keefover-Ring et al. 2009), contains up to 46 % 1,8-cineole in its essential oil.

In recent years the use of genetic and molecular methods has become popular to investigate the mechanisms involved in controlling secondary metabolite variation in plants. For example, a study on cucurbitacins, bitter C_{30} terpenoid compounds that differ among

cucumber varieties (*Cucumis sativus*), identified a genetic locus which seems to be involved in an epistatic control of these compounds (Zhang et al. 2013). Other examples involve the Myrtaceae plant family, which is well known like the Lamiaceae for a terpenoid-containing essential oil that varies among species. In *Eucalyptus nitens* and *E. globulus*, quantitative trait loci involved in the genetic control of foliar terpenes were identified (Henery et al. 2007, O'Reilly-Wapstra et al. 2011). Also in the medicinal tea tree (*Melaleuca alternifolia*), for which six monoterpene chemotypes haven been identified, a recent study analysed the transcription of enzymes involved in the basic pathways of terpene biosynthesis (the MEP and MVA pathways) and their possible role in chemotype formation (Webb et al. 2013). A few studies have also focused on the influence of ploidy levels on secondary metabolite biosynthesis in plants. One of these studies compared the mono- and sesquiterpene content of the 'goldenrod' *Solidago gigantea* (Asteraceae) plants with their ploidy level. Three ploidy levels (17 % hexaploid, 58 % tetraploid, and 25 % diploid) were encountered in the native range in the United States, while only diploid (18 %) and teraploid (82 %) cytotypes were found in the introduced range in Europe. No significant differences in mono- and sesquiterpene levels could be found between diploid and tetraploid plants on the same continent. However, the levels of secondary metabolites in Europe were lower than in North America, which was explained by the likely absence or reduced presence of specialist herbivores in the new environment (Hull-Sanders et al. 2009). A study conducted on members of the monocotyledonous Iridaceae, analysed the possible correlation of genome dosage to the carotenoid biosynthesis in the stigmata of the triploid 'saffron crocus' *Crocus sativus* and two of it's closer diploid relatives *C. cartwrightianus* and *C. hadriaticus*. The obtained results indicated that the higher apocarotenoid amount present in the stigmata of *C. sativus* is due to an interaction of carotenogenic gene dosage and their regulation genes (Castillo et al. 2005).

Concerning thyme, it has been shown that chemotypes growing in mixed populations dominated by one or two other chemotypes possess significantly higher amounts of the dominant monoterpenes of other chemotypes in their essential oil than when growing in unmixed populations (Thompson et al. 2003). These observations suggest that the genetic regulation of the chemotype formation in *T. vulgaris* is quite complex and could involve epistatic control, among other factors. Polyploidy is well-known to be an important factor in plant evolution by giving rise to novel variation and morphologies that lead to speciation. Most plants are assumed to have undergone polyploid formation during their evolution

(Osborn et al. 2003) and thus ploidy level might also be involved. *T. vulgaris* is known as a diploid plant with 2n = 28 or 2n = 30 (Elena-Rosselló 1981, Stahl-Biskup and Sáez 2002). However, several members of the genus *Thymus* have been shown to be polyploid. Thus it is possible that chemotype formation in *T. vulgaris* is in part a result of polyploidy.

This chapter describes the biochemical characterization of the essential oils from the *T. vulgaris* chemotypes (G_1, A_{10}, U_6, L_{48}, C_{10}, T_2, and T_{28}) used in this thesis by GC-MS, as well as the ploidy level determination of these chemotypes by flow cytometry.

Material and Methods

Plant material and chemicals

The *T. vulagris* chemotypes, G_1, A_{10}, U_6, L_{48}, C_{10}, T_2, and T_{28} (for abbreviations see Results), were obtained from Prof. John D. Thompson, Montpellier, France. The plants were descended from those originally collected from wild populations in southern France and propagated at the CEFE-CNRS experimental garden in Montpellier (Amiot et al. 2005, Linhart et al. 2005). The plants were grown in the greenhouse and supplemented with 13 h of artificial light (sodium lamps, 320 µmol m^{-2} s^{-1} of photosynthetic active radiation), at a temperature cycle of 20 – 22 °C / 18 – 20 °C (day / night), and 55 % relative humidity. The plants were grown in pots (14 cm in diameter) with drainage, in a composition of 63 % clay (Tonsubstrat, Klasmann Deilmann GmbH, Geeste, Germany), 9 % sand (0.7 – 1.2 mm), 13,5 % vermiculite (1 - 2 mm), and 13,5 % hydroponic clay pebbles (2 - 4 mm), watered with tap water approximately every three days, and fertilized with a complete fertilizer, 0.1 – 0.2 % Flory® 3 (Planta Düngemittel GmbH, Gegenstauf, Germany), every 4 – 6 weeks. Every three months the plants were newly propagated from stem cuttings to ensure good growth and yield. Young leaf tips including the first fully expanded leaves where used for all experiments. *T. vulgaris* variety 'Deutscher Winter' seeds (Quedlinburger Saatgut GmbH, Quedlinburg, Germany) were purchased from a local garden center and used as reference material for the ploidy determination of the chemotypes from France.

Common laboratory chemicals were bought from Carl Roth (Carl Roth GmbH & Co. KG, Karlsruhe, Germany), Sigma Aldrich Chemie (Sigma Aldrich Chemie GmbH, Taufkirchen, Germany), and Bio-Rad (Bio-Rad Laboratories, Hercules, USA) if not mentioned otherwise in this thesis.

Flow cytometry for ploidy determination

The Cell Counter Analyzer CCA-II together with the CyStain® UV Precise P-Kit (both Partec GmbH, Münster, Germany) were used to determine the ploidy levels of the plants obtained from France. Following the manufacturer's instructions, young shoots with fully expanded leaves were harvested and stored on ice. From the fully expanded leaves, the leaf tips and stems were cut off before adding 400 µl nuclei-extraction buffer. The leaves were then chopped to small pieces with a razor blade, and left at room temperature for 10 – 30 s, before filtering the solution through filters (Cell Trics 30 µm) to remove cell debris. To this filtrate, 1.6 ml ice cold DAPI staining solution were added and the mixture was analysed

in the flow cytometer. Plants of the known diploid thyme variety 'Deutscher Winter' were used as a reference.

Terpene extraction from leaves

Young leaf tips including the first fully expanded leaves from 22 plants of each chemotype were pooled and frozen in liquid N_2 immediately after harvest and ground to a fine powder in chilled mortars and pestles. The powder (50 – 100 mg) was soaked/extracted in 1 ml ethyl acetate:pentane (2:1) containing the internal standard (menthol, 50 ng/µl) for 24 h at room temperature with constant rotation. The solution was then cleared with activated charcoal for 5 min and dried over a column of 500 mg anhydrous Na_2SO_4 powder. All extractions were performed in triplicate.

GC-MS analysis of plant volatiles

The terpene extracts from thyme leaves were analysed by gas chromatography (Agilent Hewlett-Packard 6890, Agilent Technologies Inc., Santa Clara, CA, USA) coupled to a mass spectrometer (Agilent Hewlett-Packard 5973, Agilent Technologies Inc.), using helium as the carrier gas (1 ml min^{-1}). For analysis, 2 µl of ethyl acetate:pentane (2:1) extracts were injected splitless with an injector temperature of 230 °C. The terpenes were separated on a DB5-MS column (30 m length, 0.25 mm inner diameter, and 0.25 µm film (J&W Scientific, Santa Clara, CA, USA)). After 2 min at 40 °C the first ramp to 175 °C was with a temperature increase of 5 °C min^{-1}, followed by a second ramp of 90 °C min^{-1} to 250 °C, with a final hold of 3 min. The mass spectrometer had a quadrupole mass selective detector and was operated with a transfer line temperature of 230 °C, source temperature of 230 °C, quadrupole temperature of 150 °C, ionization potential of 70 eV, and scan range of *m/z* 40 - 350. All terpene products were identified by using the Agilent Technologies Inc. software with the Wiley275.L and NIST98.L MS libraries, as well as by comparison of mass spectra and retention times with those of authentic standards (Sigma-Aldrich Chemicals, Steinheim, Germany).

Results

Thymus vulgaris chemotypes show a high variability in essential oil composition

The chemotypes of *T. vulgaris* do not differ significantly in their morphological phenotypes (Figure 2-1), and have been distinguished only by olfactorial or GC-MS measurements. Seedlings obtained from southern France, were assessed for the terpene content of their essential oil and seven chemotypes were selected for this thesis. To this end the volatiles of ground fresh seedling leaves were extracted by solid micro phase extraction and analyzed by gas chromatography coupled to a mass spectrometer on a DB5-MS column, to determine which plants should be chosen for further GC-MS analysis. The chosen plants (Figure 2-2) were propagated and grown in the greenhouse under controlled conditions and their essential oils were extracted with an ethyl acetate:pentane blend (2:1) and further analyzed by GC-MS (Table 2-1).

The essential oils of *T. vulgaris* chemotypes described to date are dominated by monoterpenes and their acetates, while sesquiterpenes contribute only a small portion to the blends. Monoterpenoids and their acetates in the analyzed chemotypes amounted to 71 - 98 % of the total essential oil, while sesquiterpenes accounted for 8 - 20 % of the oil. The major terpenoid components were monoterpene alcohols (e.g. geraniol, linalool, thymol, carvacrol), monoterpene hydrocarbons (e.g. *α*-thujene, *α*- and *β*-pinene, limonene, sabinene), and sesquiterpene hydrocarbons (*E*)-*β*-caryophyllene, germacrene D, and bicyclogermacrene, as well as the acetates of several of the monoterpene alcohols.

The terpenoid blend of the geraniol chemotype 'G_1', used in this thesis, is an average representative of the geraniol chemotype found in France and contained about 5 % neral, 11 % geraniol, 16 % geranial, and 43 % geranyl acetate, as well as 15 % (*E*)-*β*-caryophyllene and 5 % germacrene D (Table 2-1 and Figure 2-2). In general it can be said that the variation in the analysis was higher than that in the other chemotypes. The distinguishing monoterpenoids found in the *α*-terpineol 'A_{10}' chemotype, *α*-terpineol and *α*-terpinyl acetate, amounted to 42 % and 22 % respectively, followed by sabinene (8 %) and myrcene (5 %), while the main sesquiterpene was (*E*)-*β*-caryophyllene with 8 %. The essential oil of the thujanol chemotype, which is also sometimes referred to as *trans*-sabinene hydrate chemotype, is traditionally considered to be a mixture of terpinene-4-ol, 4-thujanol (*trans*-sabinene hydrate), 8-myrcenol, and linalool (Thompson et al. 2003). The essential oil of the thujanol chemotype used in this thesis, the 'U_6' chemotype, is indeed a mixture of several

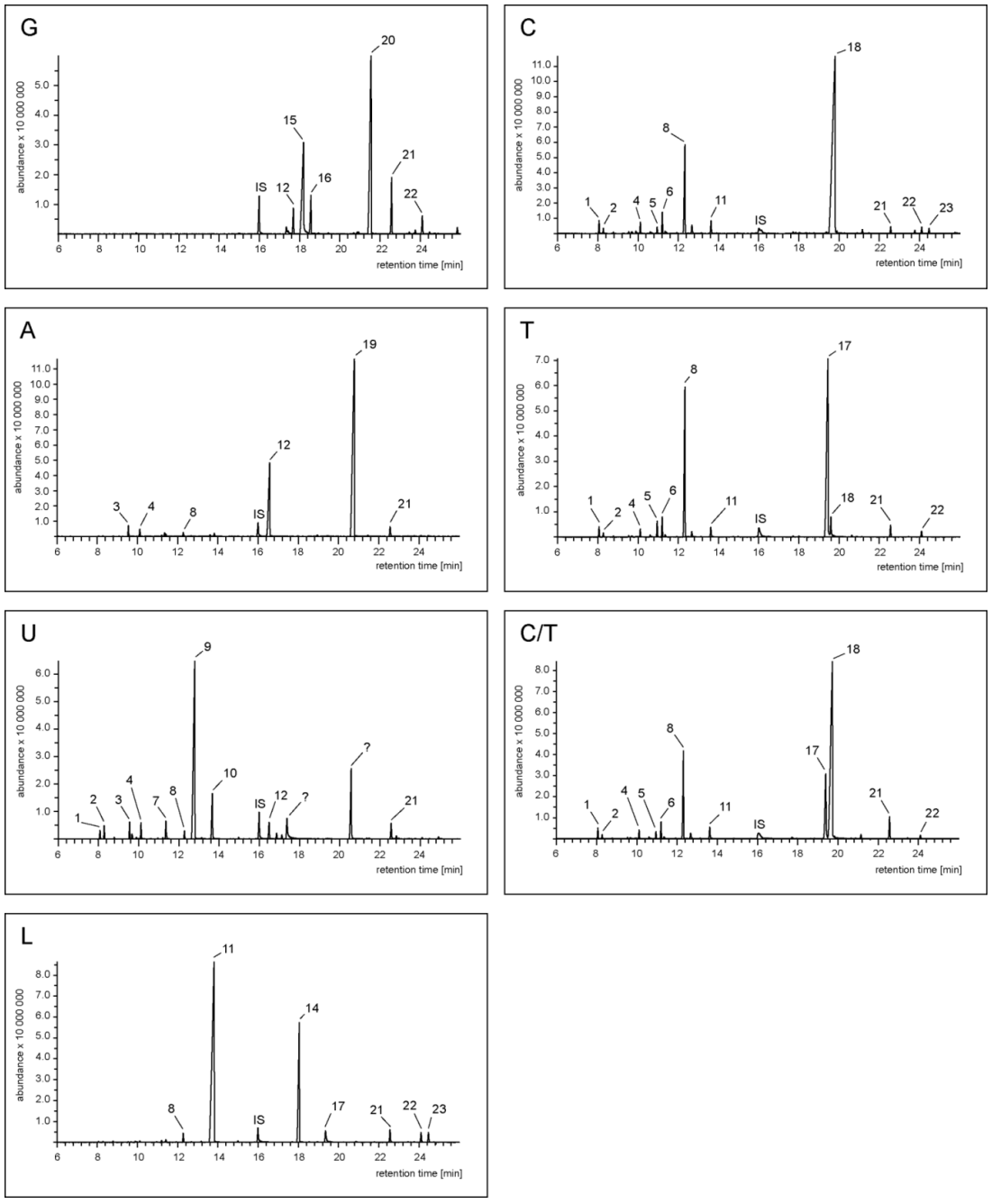


Figure 2-2: Terpene composition of the *T. vulgaris* chemotypes studied in this thesis. Mono- and sesquiterpene volatiles were extracted from ground fresh leaves by solid phase micro extraction (SPME) and analyzed by GC-MS. The compounds were identified as: (1) *α*-thujene, (2) *α*-pinene, (3) sabinene, (4) myrcene, (5) *α*-terpinene, (6) *p*-cymene, (7) limonene, (8) *γ*-terpinene, (9) *trans*-sabinene hydrate, (10) *cis*-sabinene hydrate, (11) linalool, (12) *α*-terpineol, (13) neral, (14) linalyl acetate, (15) geraniol, (16) geranial, (17) thymol, (18) carvacrol, (19) *α*-terpinyl acetate, (20) geranyl acetate, (21) (*E*)-*β*-caryophyllene, (22) germacrene D, (23) bicyclogermacrene, (IS) menthol.

monoterpenoids, none of which dominate the oil as much as the denominating monoterpenoids of the other chemotypes. The two sabinene hydrates, *trans*- and *cis*-sabinene hydrate, make up 30 % of the essential oil, followed by 7 % each of limonene and *α*-terpineol, as well as noteworthy amounts of myrcene (6 %), sabinene (6 %), and *α*-pinene (4 %), while (*E*)-*β*-caryophyllene (7 %) is the main sesquiterpene. In addition to this is an unidentified terpenoid, which amounts up to 20 % of the essential oil. The linalool chemotype

Chemotype / Terpene	**G_1**	**A_{10}**	**U_6**	**L_{48}**	**C_{10}**	**T_2**	**T_{28}**
α-thujene	0.00 (0.00)	0.07 (0.07)	2.78 (0.05)	0.21 (0.11)	5.17 (0.61)	4.44 (0.41)	3.60 (0.71)
α-pinene	0.00 (0.00)	0.15 (0.10)	4.34 (0.10)	0.15 (0.12)	2.06 (0.24)	1.84 (0.20)	1.41 (0.28)
camphene	0.03 (0.03)	0.15 (0.10)	0.46 (0.04)	0.17 (0.17)	0.48 (0.08)	0.48 (0.13)	0.16 (0.08)
sabinene	0.00 (0.00)	8.10 (2.15)	5.82 (0.16)	0.05 (0.05)	0.70 (0.12)	0.48 (0.13)	0.42 (0.08)
β-pinene	0.00 (0.00)	0.09 (0.06)	1.43 (0.08)	0.03 (0.03)	0.55 (0.11)	0.39 (0.06)	0.31 (0.08)
myrcene	0.30 (0.30)	5.26 (1.62)	6.29 (0.44)	0.55 (0.16)	4.75 (1.44)	3.29 (0.43)	2.80 (0.72)
α-phellandrene	0.00 (0.00)	0.00 (0.00)	0.13 (0.13)	0.00 (0.00)	0.76 (0.13)	0.82 (0.08)	0.55 (0.08)
α-terpinene	0.00 (0.00)	(0.06) 0.06	(0.68) 0.44	0.31 (0.03)	4.74 (0.23)	7.71 (0.38)	3.69 (0.28)
p-cymene	0.02 (0.02)	0.40 (0.35)	0.19 (0.10)	0.65 (0.17)	6.69 (0.81)	12.76 (4.92)	5.48 (1.45)
limonene	0.18 (0.18)	3.02 (0.73)	7.12 (0.28)	0.01 (0.01)	1.27 (0.36)	1.43 (0.23)	0.94 (0.19)
1,8-cineole	0.00 (0.00)	2.01 (0.48)	0.00 (0.00)	1.02 (0.18)	0.58 (0.11)	0.00 (0.00)	0.00 (0.00)
β-ocimene	0.30 (0.30)	0.00 (0.00)	0.00 (0.00)	0.13 (0.08)	0.35 (0.14)	0.14 (0.07)	0.13 (0.07)
γ-terpinene	0.00 (0.00)	2.75 (0.93)	3.09 (0.69)	4.00 (0.56)	26.62 (9.24)	21.71 (12.62)	23.33 (9.10)
cis-sabinene hydrate	0.00 (0.00)	0.00 (0.00)	18.03 (2.39)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
linalool	0.00 (0.00)	1.04 (0.56)	0.00 (0.00)	26.91 (4.44)	5.04 (0.86)	4.03 (2.00)	5.34 (1.36)
trans-sabinene hydrate	0.00 (0.00)	0.26 (0.26)	11.80 (1.15)	0.09 (0.09)	3.63 (0.51)	2.76 (0.41)	2.64 (0.78)
α-terpineol	0.00 (0.00)	42.15 (12.15)	7.20 (0.49)	0.09 (0.09)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
α-terpinyl acetate	0.00 (0.00)	21.89 (1.48)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Chemotype / Terpene	G_1	A_{10}	U_6	L_{48}	C_{10}	T_2	T_{28}
geraniol	34.18 (10.84)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
geranyl acetate	43.35 (13.99)	0.68 (0.43)	0.11 (0.11)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
linalyl acetate	0.00 (0.00)	0.15 (0.15)	0.00 (0.00)	37.6 (12.86)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
unknown	0.00 (0.00)	0.00 (0.00)	20.92 (3.57)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
thymol	0.00 (0.00)	0.07 (0.07)	0.34 (0.17)	9.52 (1.75)	0.86 (0.07)	15.19 (2.62)	20.33 (7.12)
thymyl acetate	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.46 (0.23)	0.08 (0.08)
carvacrol	0.00 (0.00)	0.28 (0.28)	0.00 (0.00)	0.33 (0.33)	17.50 (5.47)	12.22 (3.15)	16.64 (6.47)
carvacryl acetate	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	2.91 (0.99)	0.00 (0.00)	1.42 (0.51)
(*E*)-*β*-caryophyllene	14.73 (4.78)	8.41 (3.06)	7.27 (0.87)	6.58 (1.41)	4.35 (1.93)	4.93 (0.97)	8.32 (2.76)
germacrene D	4.75 (1.68)	0.57 (0.37)	0.61 (0.06)	5.15 (1.07)	4.30 (1.92)	2.75 (0.37)	1.53 (0.34)
bicyclogermacrene	0.51 (0.15)	0.45 (0.24)	0.08 (0.06)	5.28 (1.05)	3.41 (1.48)	0.00 (0.00)	0.00 (0.00)

Table 2-1: Essential oil composition of the *T. vulgaris* chemotypes analyzed in this thesis in percent. Standard error (± SE) is shown in parenthesis (n = 3). G_1 – geraniol chemotype; A_{10} – *α*-terpineol chemotype; U_6 – *trans*-sabinene hydrate chemotype; L_{48} – linalool synthase; C_{10} – carvacrol chemotype; T_2 – carvacrol/thymol chemotype; T_{28} – carvacrol/thymol chemotype.

'L_{48}' contained 27 % linalool and 38 % linalyl acetate in its essential oil, next to 7 % (*E*)-*β*-caryophyllene, 5 % germacrene D, and 5 % bicyclogermarene, as well as noteworthy amounts of *γ*-terpinene (4 %) and thymol (10 %).

In contrast to the non-phenolic chemotypes, with the exception of the U chemotype, the essential oil blend of the phenolic chemotypes is quite diverse in noteworthy terpenoids. Thus the carvacrol 'C_{10}' chemotype essential oil contains the major monoterpenoid constituents *γ*-terpinene (27 %), carvacrol (18 %), and *p*-cymene (7 %), and also 5 % of each *α*-terpinene, *α*-thujene, myrcene, and linalool, and 4 % of each (*E*)-*β*-caryophyllene and germacrene D. The two chemotypes 'T_2' and 'T_{28}' which had been chosen as representatives of a thymol and a carvacrol/thymol chemotype based on the preliminary SPME analysis (Figure 2-2), turned out to be both mixed carvacrol/thymol chemotypes. The terpenoid blend of the 'T_2' chemotype was found to be dominated by *γ*-terpinene (22 %) and the phenolic

monoterpenes *p*-cymene (13 %), thymol (15 %), and carvacrol (12 %), while containing also noteworthy amounts of *α*-terpinene (8 %), *α*-thujene (4 %), and linalool (4 %), as well as small amounts of the sesquiterpenes (*E*)-*β*-caryophyllene (5 %) and germacrene D (3 %). The essential oil blend found in the 'T_{28}' chemotype was quite similar, with 23 % *γ*-terpinene, 20 % thymol, 17 % carvacrol, and 5 % linalool, while *p*-cymene and *α*-terpinene amounted to only 5 % and 4 %, respectively. (*E*)-*β*-Caryophyllene made up about 8 % of the oil while only minor amounts of germacrene D (2 %) could be detected.

The ploidy level of plants can have an influence on secondary metabolism, as has been shown for example in different *Crocus* species (Castillo et al. 2005). It was therefore necessary to determine the level of ploidy of the different monoterpene chemotypes of *T. vulgaris* that were investigated in this thesis. Young shoots with fully expanded leaves were used for the flow cytometry and plants of the known diploid *T. vulgaris* variety 'Deutscher Winter' were utilized as a reference. The results showed that all analyzed chemotypes were diploids.

Discussion

Chemical polymorphism is a phenomenon which is neither restricted to a certain class of secondary metabolites nor to a specific plant family. However, the intraspecific variation of the terpenoid composition of the essential oils of the Lamiaceae is especially well-known and has been best investigated in *T. vulgaris*. Seven distinct monoterpenoid (geraniol, *α*-terpineol, thujanol-4, linalool, carvacrol, thymol, and 1,8-cineole) chemotypes have been described for this species, while a close relative, *T. praecox* subsp. *arcticus* is known to possess 17 distinct chemotypes, several of which are sesquiterpene chemotypes (Schmidt et al. 2004). Several of the described chemotypes occur in both species. For example, both species possess an *α*-terpineol chemotype which contains both *α*-terpineol and *α*-terpinyl acetate as the main essential oil compounds. On the other hand, two linalool chemotypes have been described in *T. praecox*, one of which contains only linalool and one of which contains linalool and linalyl acetate, while only the latter can be found in *T. vulgaris*.

While some early studies on the essential oil of *T. vulgaris* chemotypes stated that these rarely contained 'foreign' monoterpenoids, meaning monoterpenoids characteristic of other chemotypes, Thompson et al. (2003), could demonstrate that such occurrences were more frequent than hitherto believed by taking the population structure and the plant collection site into consideration (Granger and Passet 1973, Simeon de Bouchberg et al. 1976). Thus it was necessary for this thesis to determine the precise essential oil composition of the plants utilized. Plants of the geraniol chemotype, for example, derived from populations which are dominated by other chemotypes, such as for instance a population dominated by the linalool chemotype, have a relatively high proportion (> 20 %) of linalool in their essential oil, while other 'foreign' monoterpenoids, such as the phenolic monoterpenes, are considerably lower and range between 5 – 10 %. Geraniol plants from sites where geraniol is the dominant chemotype in the population, on the other hand contain only minor amounts of 'foreign' monoterpenes (Thompson et al. 2003). In this study, the lack of significant amounts of 'foreign' monoterpenoids, such as linalool, found in the essential oil of the geraniol chemotype G_1 analyzed, makes this plant most likely one from a population where geraniol is the dominant chemotype. However the overall amount of geraniol and its acetate (78 %) is less than the range of 81 – 95 % found in the literature, while the sesquiterpenes (*E*)-*β*-caryophyllene and germacrene D contribute almost 20 % to the essential

oil (Table 2-1)(Passet 1971, Simeon de Bouchberg et al. 1976, Thompson et al. 2003). The (*E*)-*β*-caryophyllene detected in thyme essential oils, normally amounts to 5 - 10 % (Thompson et al. 2003), while it is 15 % in this chemotype, the highest percentage among all tested chemotypes in this thesis, and otherwise ranges between 4 % (C_{10}) and 8 % (A_{10} and T_{28}). The germacrene D amount found in the G_1 chemotype is equal to the one detected in the linalool L_{48} chemotype (5 %), while the other tested chemotypes range between zero (T_2 and T_{28}) and 4 % (C_{10}).

Similar to the G_1 geraniol chemotype the non-phenolic *α*-terpineol chemotype A_{10} also likely comes from a population dominated by the *α*-terpineol chemotype as its essential oil contains predominately *α*-terpineol and *α*-terpinyl acetate (64 %), while the distinguishing monoterpenoids from other chemotypes are present in only minor amounts. The essential oil composition of the *α*-terpineol chemotype is, according to the literature, dominated by 75 – 96 % *α*-terpineol and its acetate ester, while small amounts of linalool and terpinen-4-ol are sometimes encountered (Passet 1971, Granger and Passet 1973, Simeon de Bouchberg et al. 1976). Next to the 64 % distinguishing monoterpenoids, there were also notable amounts of sabinene (8 %), myrcene (5 %), and limonene (3 %) found, which comprised, together with the sesquiterpenes 26 % of the A_{10} chemotype oil (Table 2-1).

In contrast to the other non-phenolic chemotypes, the thujanol chemotype U is known for its mélange of several monoterpenoids and its lack of a predominant monoterpenoid. *Trans*- and *cis*-sabinene hydrate, which appear to be never esterified, can account for up to 56 % of the oil, while there are also significant amounts of *α*-terpineol, linalool, 8-myrcenol, and terpinen-4-ol, together with their esters, some of which can amount to more than 15 % (Passet 1971, Simeon de Bouchberg et al. 1976, Thompson et al. 2003). The analyzed U_6 chemotype appears to be, similar to the two aforementioned chemotypes, from a site where the U chemotype is the major one, since its oil is dominated by *cis*- and *trans*-sabinene hydrate (30 %) together with several other monoterpenoids, such as myrcene (6 %), limonene (7 %), and sabinene (6 %), while the only 'foreign' monoterpenoid occurring in noteworthy amounts, *α*-terpineol (7 %), is well within the margins stated in the literature (Table 2-1).

The remaining chemotypes (L_{48}, C_{10}, T_2, and T_{28}), are in comparison to this most likely derived from sites where these are not the dominant chemotypes, since their essential oil composition is not as distinct as expected. The essential oil of the linalool chemotype, for instance, is described to be comprised of 85 – 95 % linalool and linalyl acetate with only

minor amounts of *trans*-sabinene hydrate (4-thuyanol), terpinen-4-ol, *α*-terpineol, thymol, or carvacrol. The oil of the linalool chemotype analyzed in this thesis however, contains relatively high amounts of *γ*-terpinene (4 %) and thymol (10 %), most likely indicating a recent thymol chemotype crossing in its ancestry. A peculiarity of the linalool chemotype is that seedlings have been described to exhibit the essential oil characteristics of a phenolic thyme chemotype in the first three months after germination, which later changes irreversibly to the linalool chemotype (Passet 1971, Vernet et al. 1986). It has been postulated that this change in essential oil composition serves as a defense strategy against herbivory. The snail *Helix aspersa*, for example, is often encountered in less xeric areas of the thyme habitat and shows the highest preference for plants of the linalool chemotype, which in turn also prefers less xeric areas, while it shuns plants of the phenolic chemotypes (Linhart and Thompson 1995, Linhart et al. 2005).

The three phenolic chemotypes, C_{10}, T_2, and T_{28}, used in this study, were not as typical as had been described in the literature. The carvacrol chemotype typically possesses an amount of 57 – 85 % carvacrol in its essential oil, with only small amounts of thymol, linalool, *trans*-sabinene hydrate, terpinen-4-ol, *α*-terpineol, and other monoterpenoids, while *p*-cymene and *γ*-terpinene can contribute up to 20 % to the oil (Passet 1971, Granger and Passet 1973, Simeon de Bouchberg et al. 1976, Thompson et al. 2003). In the case of the C_{10} chemotype, the essential oil contains only 18 % carvacrol, while the remainder is very diverse and is comprised of *γ*-terpinene (27 %), *p*-cymene (7 %), *α*-terpinene (5 %), *α*-thujene (5 %), linalool (5 %), myrcene (5 %), *trans*-sabinene-hydrate (4 %), as well as the ubiquitous (*E*)-*β*-caryophyllene (4 %) and germacrene D (4 %). The two remaining chemotypes, T_2 and T_{28}, possess comparable amounts of carvacrol (12 and 17 %, respectively) and thymol (15 and 20 %, respectively), making them most likely heterozygous forms of the carvacrol chemotype. The essential oil of the thymol chemotype can in general contain up to 27 % carvacrol next to thymol, while other monoterpenes such as linalool, *trans*-sabinene hydrate, or *α*-terpineol can also contribute considerable amounts to the oil, which makes it an extremely variable chemotype (Passet 1971, Thompson et al. 2003). This chemotype is also known to form less dense and homogenous populations than the carvacrol chemotype, and individuals of pure thymol chemotypes appear less frequently than individuals of pure carvacrol chemotypes (Passet 1971, Simeon de Bouchberg et al. 1976). Similar to the C_{10} chemotype, *γ*-terpinene is the main essential oil component in T_2 (22 %) and T_{28} (23 %), while other monoterpenes such

as *p*-cymene (13 and 6 %, respectively), *α*-terpinene (8 and 4 %, respectively), or linalool (4 and 5 %, respectively) too can be found as components of the essential oil (Table 2-1).

It should also be noted, that the 1,8-cineole chemotype ('E'), which was first characterized in Spain and has been more recently identified in the Saint-Martin-de-Londres basin in southern France, was not included in this thesis (Adzet et al. 1977, Keefover-Ring et al. 2009).

It has already been stated that ploidy can have a significant influence on the plant secondary metabolism, although studies focusing on terpenoid metabolism in particular are sparse. In the early studies of Passet and Granger, the level of ploidy of certain chemotypes (linalool, *α*-terpineol, and carvacrol) was determined to be 2n = 30 (Passet 1971, Granger and Passet 1973). We obtained similar results for all chemotypes analyzed in this thesis by using flow cytometry with the known diploid variety 'Deutscher Winter' as a reference. This indicates that the regulation mechanism underlying the monoterpene polymorphism in *T. vulgaris* is not a result of deviating ploidy levels but is in fact controlled in a different manner.

3 Characterization of terpene biosynthetic genes of *Thymus vulgaris*

Introduction

The principal constituents of essential oils in the Lamiaceae, such as basil, oregano, mint, and thyme, are mono- and sesquiterpenes. These compounds are synthesized and stored in glandular hairs, situated on the aerial parts of the plants, and are volatilized when the tissue is crushed or damaged. The biosynthesis of terpenoids, in general, can be divided into four stages: (i) the synthesis of the biological five-carbon isoprene unit, (ii) repetitive condensations of the five-carbon unit to form a series of larger and larger prenyl diphosphates, (iii) conversion of prenyl diphosphates to the basic terpenoid skeletons, and (iv) further modifications to the basic skeletons, including oxidation, reduction, isomerization, conjugation, and other transformations.

While monoterpene biosynthesis starts with the ten-carbon unit geranyl pyrophosphate (GPP), the sesquiterpene biosynthesis starts with the fifteen-carbon unit farnesyl pyrophosphate (FPP). These acyclic prenyl diphosphates serve as the precursors for the monoterpene synthases and sesquiterpene synthases, respectively, and can, after conversion to basic terpenoids, be further modified by other enzymes, such as for example cytochrome P450 enzymes. Terpene synthases in general employ a carbocationic reaction mechanism similar to that of prenyltransferases. The reaction is initiated by ionization of the diphosphate group which requires a divalent metal ion, such as Mg^{2+} or Mn^{2+}. Thus the enzyme-bound, allylic carbocation can be cyclized by the addition of the resonance-stabilized cationic center to one of the other carbon-carbon double bonds in the substrate. Cyclization can be then followed by a series of rearrangements, including hydride and alkyl shifts, and additional cylizations, mediated through enzyme-bound carbocationic and neutral intermediates. The multiple fates of these reactive carbocations are responsible for the great diversity of terpene synthases products. For example, the γ-humulene and δ-selinen synthase from grand fir (*Abies grandis*) produce more than 50 different products, while the linalool synthase from bergamot mint (*Mentha citrata*) produces essentially only a single terpenoid product (Steele et al. 1998, Hyatt and Croteau 2005, Tholl et al. 2005, Chen et al. 2011). The

reaction cascade of the terpene synthases is eventually terminated by deprotonation of the cation to form a new double bond or by the capture by a nucleophile, such as water.

Several representatives of the Lamiaceae, such as mint (Croteau et al. 2005), basil (Iijima et al. 2004a, Iijima et al. 2004b), sage (Wise et al. 1998, Hoelscher et al. 2003, Kampranis et al. 2007), or perilla (Hosoi et al. 2004, Masumoto et al. 2010), have been well studied as far as the biosynthesis of their main monoterpene constituents are concerned. It can therefore be deduced that the dominant non-phenolic monterpenes, such as geraniol, *α*-terpineol, sabinene hydrate, and linalool, as well as the phenolic monoterpene precursor, *γ*-terpinene, are likely to be direct products of monoterpene synthases, rather than formed by another mode of biosynthesis. Similar assumptions can be made for the two main sesquiterpenes, germacrene D and (*E*)-*β*-caryophyllene, found in the essential oil of *T. vulgaris*. The formation of the two dominant phenolic monoterpenes thymol and carvacrol, on the other hand, most likely involves the help of another enzyme class, the cytochrome P450 enzymes (P450). Members of this large family of membrane-bound catalysts are known to use molecular oxygen and NADPH to catalyze the position-specific oxidation of many terpenoid substrates. Menthol biosynthesis in peppermint, for example, involves several redox enzymes and an isomerase to transform limonene to menthol (Mahmoud and Croteau 2003). It is most likely that the formation of thymol and carvacrol also require the enzymatic activity of P450s. Poulose and Croteau (1978a) suggested that the olefin *γ*-terpinene is the precursor to these two dominant phenolic monoterpenes in *T. vulgaris*.

In this chapter we describe the isolation and identification of three monoterpene synthases, one sesquiterpene synthases and two P450 hydroxylases responsible for the production of major essential oil components in *T. vulgaris*, as well as the partial isolation of two monoterpene synthases and a sesquiterpene synthase also thought to be of importance to the essential oil formation in thyme.

Material and Methods

Isolation of secretory cell clusters from thyme

The isolation of glandular trichomes followed a modified method first described for mint by Gershenzon et al. (1992). Young leaves (apical buds and newly expanded leaves from the first upper node of vegetative growing shoots) were harvested before midday and soaked in ice-cold *a. dest.* containing 0.05 % [v/v] Tween20™ for 1 - 2 h on ice before being rinsed twice with sterile water without Tween20™. Glandular trichomes were abraded in a finned 350 ml polycarbonate chamber (BeadBeater, Biospec Products, Bartlesville, USA) filled with 10 – 15 g plant material, 65 ml glass beads (0.5 – 1.0 mm diameter), Amberlite® XAD-4 resin (1.5 g g^{-1} plant material), and ice-cold extraction buffer (25 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) adjusted to pH 7.3 with KOH, containing 200 mM D-sorbitol, 12 mM KCl, 10 mM sucrose, 5 mM $MgCl_2$, 5 mM DTT (dithiotreitol), 0.5 mM K_2HPO_4, 0.1 mM $Na_4P_2O_7$, 1 % [w/v] PVP 40 (polyvinylpyrrolidone, M. 40.000), 0.6 % [w/v] methyl cellulose (~ 25 ctpoise)). Leaves and buds were abraded with 3 pulses of operation for 1 min at low to medium speed. The whole isolation procedure was carried out at 4 °C, and after each pulse the chamber was allowed to cool on ice for 1 min. To remove the plant material, glass beads, and XAD-4 resin, the chamber content was filtered through a 500 µm nylon mesh (Sefar Printing Solutions Inc., Lumberton, USA). The remanent plant material and beads were scraped from the mesh and rinsed once with additional extraction buffer that was also passed through the 500 µm mesh. The 500 µm filtrate was then passed through a series of nylon meshes with decreasing mesh sizes (350 µm, 200 µm, and 100 µm) to remove small tissue fragments. Finally, it was possible to collect the secretory cell clusters by passing the 100 µm mesh filtrate through a 20 µm mesh, due to their average diameter of 60 µm. The isolation buffer used in these filtering steps was extraction buffer that lacked PVP 40 and methyl cellulose, in order to lower the viscosity. Collected cell clusters were resuspended in isolation buffer and checked for cluster concentration with a microscope (Axiovert 200, Carl Zeiss, Jena, Germany) before they were frozen in liquid nitrogen and stored at -80 °C. Specific information on the thyme plants and their growing conditions are described in Chapter I.

Isolation of total RNA from thyme

Total RNA extraction from thyme was accomplished by using a method based on Trizol® reagent, a monophasic solution of phenol and guanidine isothiocyanate. Depending

on the designated usage of the isolated RNA, either isolated glandular trichomes or whole young leafs where used as source material (see Chapter I).

Homogenized tissue was mixed with 1.5 ml ice-cold Trizol® reagent and briefly vortexed before being incubated for 3 min on ice. 200 µl of ice-cold chloroform were added to the tube, vortexed, and incubated for another 2 min on ice. The sample was then centrifuged at 4 °C for 15 min at 10.500 rpm; and the supernatant was collected in a new tube before spun again for 5 min at 13.000 rpm and 4 °C. If the supernatant was still cloudy another 200 µl of ice-cold chloroform were added and the centrifugation repeated. To the clear supernatant 500 µl isopropanol were added and incubated for 10 min at room temperature. After spinning the sample for 10 min at 10.500 rpm and 4 °C, the supernatant was taken off and the pelleted RNA was washed with 1 ml ice-cold 75 % EtOH. After centrifugating the sample for 5 min at 8000 rpm at 4 °C, the supernatant was discarded and the pellet dried for a maxium of 5 min before the pellet was resuspended in 100 µl RNase-free H_2O. RNA designated for cDNA library construction was further purified by using the RNeasy Mini Kit from Qiagen (Hilden, Germany), following the manufacturer's instructions for RNA cleanup.

cDNA library construction and colony lifts

Two different cDNA libraries were constructed for this thesis. SMART™ RACE cDNA libraries (5'- and 3'-cDNA) were prepared for all seven chemotypes, and Creator™ SMART™ cDNA libraries were constructed from the chemotypes T_{28} and A_{10}.

SMART™ RACE cDNA library construction

For the first-strand synthesis reaction 1 µg total RNA was used as template with the SMART™ RACE cDNA amplification kit (BD Bioscience Clontech, Mountain View, CA, USA) following the manufacturer's protocol, safe for using Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) to perform the first strand synthesis. The first-strand reaction product was then diluted in 100 µl Tricine EDTA buffer and stored at – 20 °C

Creator™ SMART™ cDNA library construction

To construct the directional Creator™ SMART™ cDNA library 1 µg total RNA was used as template with the Creator™ SMART™ cDNA library construction kit (BD Bioscience Clontech, Mountain View, CA, USA) following the manufacturer's protocol, save for using Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) to perform the

first strand synthesis, and performing the size fractionation with SizeSep^{TM}400 Spun Columns (Amersham Biosciences, Uppsala, Sweden). For the electro-transformation *E. coli* ElectroMAX DH5α were thawed on ice and 5 µl desalted DNA were added to 25 µl cell suspension and immediately transferred to chilled (-20 °C) electro-transformation cuvettes (Molecular BioProducts, San Diego, USA). Cells were pulsed at voltage: 2.0 kV, resistance: 200 Ω, and capacitance: 20 µF with a Bio-Rad *E.coli* Pulser. The electroporated cells were taken up in 970 µl SOC medium (room temperature) and incubated for 1 h at 37 °C and 220 rpm, before being spread on prewarmed LB-agar plates containing selection medium. The library was then amplified as described in the manufacturer's protocol and stored at -80 °C until further use.

Colony lifts of Creator™ SMART™ cDNA library

To prepare colony lifts of the Creator^{TM} SMART^{TM} cDNA library, agar plates with the desired colony density where chilled at 4 °C for 30 min to facilitate the procedure. Nylon Membranes for Colony and Plaque Hybridization (Roche Diagnostics Corporation, Indianapolis, USA) where placed on the agar and left there for 1 – 2 min. The membrane was then placed on 2 ml denaturation solution (0.5 M NaOH, 1.5 M NaCl) and left for 15 min at room temperature. After briefly drying the membrane on Whatman 3 MM paper it was transferred on 2 ml neutralization solution (1 M Tris-HCl pH 7.4, 1.5 M NaCl) and left to incubate for another 15 min. The membrane was briefly dried and placed on 2 ml 2x SSC solution for 10 min before the transferred DNA was crosslinked in an UV-Stratalinker® 2400 (Stratagene, La Jolla, USA) two times at 120 mJ $cm^{2\,-1}$. To remove cell debris, the membrane was treated with proteinase K (2 mg ml^{-1} solubilised in 2x SSC) for 1 h at 37 °C. Whatman 3 MM paper was moistened with *a. dest.* and then placed atop the membrane and firmly pressed down to facilitate the removal of the proteinase-treated cell debris, and then lifted off the disc. After the cell debris had been removed, the filter was ready for hybridization with radioactive labelled probes (see Chapter III).

PCR, restriction, cloning, and bacterial strains

PCR-Parameters

The components of the PCR reaction were 0.8 µl Advantage Taq DNA Polymerase Mix (5 U / µl), 5 µl 10 x Advantage Taq PCR-buffer, 1 µl dNTPs (10 mM each), 5 µl universal primer mix and 1 µl gene specific primer (10 pmol / µl), 0.5 – 1 µl cDNA, and PCR-grade water added to a final volume of 50 µl. The PCR was conducted with an initial

denaturation at 96 °C for 2 min, 30 – 35 cycles of denaturation at 96 °C for 30 sec, annealing ranging from 50 - 58 °C for 30 sec, extension at 68 °C for 60 - 150 sec, and a final step at 68 °C for 5 min. The PCR fragments were analyzed by cloning into pCR4-TOPO vector, and subsequently sequenced. The oligonucleotide primers used in this study were purchased from MWG-Biotech (Ebersbach, Germany) and Invitrogen, and were among others designed by utilizing sequence similarities to terpene synthase genes from other Lamiaceae (Supplementary Material, Table S1). For the isolation of terpene synthase genes 5'- and 3'-RACE libraries of the linalool (L_{48}), the *trans*-sabinene hydrate (thujanol) (U_6), and the carvacrol/thymol (T_{28}) chemotype were used.

DNA-agarose gel electrophoresis

The separation of DNA according to fragment size was performed with electrophoresis in 1.2 % agarose gels (1.2 % [w/v] agarose, 1 µg ml^{-1} EtBr, 0.5 x TAE buffer: 20 mM Tris acetate pH 8.2, 1 mM Na-EDTA). The DNA samples were mixed with 1/6 volume loading buffer (50 % [v/v] glycerol, 0.05 % [w/v] bromophenol blue, 0.04 % [w/v] xylene cyanol, 100 mM Na-EDTA pH 8.0) and applied to the gel slots. To compare DNA fragments' lengths, a 1 kb marker or a 10 kb marker (both Invitrogen, Carlsbad, USA) were separated together with DNA samples. Electrophoresis was performed in 0.5 x TAE buffer in an *i*-Mupid mini electrophoresis unit (Eurogentec, Seraing, Belgium) or in an EC360M Maxiwell system (EC Apparatus Corporation, St. Petersburg, USA) at constant voltages of 80 – 135 V. For visualization and quantification the gel documentation system GeneGenius and the software GeneSnap and GeneTools from Syngene (Cambridge, UK) were used.

To extract DNA from agarose gels or to purify PCR products from PCR reaction solutions, the NuceloSpin Extract kit from Macherey-Nagel (Düren, Germany) was used following the manufacturer's instructions. To promote the binding of the DNA to the silica membrane of the column, the agarose pieces were dissolved in buffer containing chaotropic salts. Washing steps with EtOH-containing buffers were followed by the elution of pure DNA with 5 mM Tris-HCl pH 8.5.

Restriction of gDNA, plasmids and PCR products

Restriction endonucleases were used to produce compatible inserts and vectors for ligation, to screen vectors for successful insertion of PCR products, and to digest genomic DNA for Southern analysis. Enzymes and the appropriate buffers (10x) were purchased from NEB (Schwalbach, Germany) or Invitrogen. All applications other than for Southern Blot

analysis (see Chapter III) used 2 – 20 U enzyme with 0.5 – 3 µg DNA and the appropriated amount of buffer in reaction volumes of 20 µl for 2 – 3 h. The applied reaction conditions were chosen to avoid possible non-specific activity of the enzymes.

Ligation and transformation

For sequencing, PCR products were ligated into the pCR4-TOPO vector following the manufacturers protocol. Ligation in other vector-systems was performed with 1 U of T4-DNA-ligase, an ATP containing buffer, and a reaction volume of 30 - 50 µl using linearized vectors and compatible PCR products. Vector and insert were ligated in an estimated molecule ratio of 1:10 and incubated at 16 °C for 5 h or overnight. Vectors used in this thesis are listed in Table S2 in the Supplemental Material section.

Competent cells were transformed with the help of a heat-shock, 50 – 500 ng DNA were added to the bacteria and left for 30 min on ice before heated to 42 °C for 32 sec and then cooled on ice for 2 min. 200 µl SOC-medium (room temperature) were added and the cells cultivated at 37 °C for 1 h before being spread on prewarmed LB-agar plates containing a suitable selection medium.

Bacterial strains and cultivation

The *E. coli*-strains were propagated submersed in liquid LB culture (25 g 'Miller's LB broth base®' Luria-Bertani medium [Gibco BRL, Carlsbad, USA] L^{-1} *a. dest.*) at 37 °C and 220 rpm in a shaker or grown on LB-agar plates (32 g 'LB agar Lennox powder' [GibcoBRL, Carlsbad, USA] L^{-1} *a. dest.*) in an incubator at 37 °C. The selective media contained antibiotics in the following concentrations: ampicillin 100 µg ml^{-1}, gentamycin 50 µg ml^{-1}, and kanamycin 50 µg ml^{-1}. Precultivation of newly-transformed bacteria was performed in SOC medium (2 g L^{-1} bacto tryptone, 0.5 g L^{-1} yeast extract, 0.5 g L^{-1} glucose, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$) for 1 h at 37 °C without antibiotics. Stock cultures were stored with 25 % [v/v] glycerol at -80 °C. Bacterial strains used in this thesis are listed in Table S3 in Supplemental Material.

Gateway® Cloning

The Gateway® Cloning system (Invitrogen) is a cloning technique developed by Invitrogen, which facilitates the transfer of heterologous DNA sequences between vectors while maintaining the reading frame. It is based on the bacteriophage lambda site-specific recombination system which facilitates the integration of lambda into the *E. coli* chromoseome and the switch between the lytic and lysogenic pathways. The system utilizes

the site-specific attachment (*att*) sites used in lambda recombination, which serve as the binding sites for recombination proteins of the lambda phage. These enzymes bring together the target sites (*att*B x *att*P and *att*L x *att*R), cleave them, and covalently attach the DNA.

The BP reaction allows the recombination of an *att*B substrate (PCR product with *att*B site or a linearized *att*B expression clone) with an *att*P substrate (donor vector with *att*P site) to create an *att*L-site containing entry clone. The thus created entry clone can be used as *att*L substrate with an *att*R substrate (desigantion vector containing an *att*R site) to create an *att*B-site containing expression clone. This LR reaction is catalyzed by the enzymes used in the lytic pathway, while the BP reaction is catalyzed by the enzymes used in the lysogenic pathway.

Heterologous expression and purification of recombinant terpene synthases

Expression constructs of the putative terpene synthases *Tv*TPS1, *Tv*TPS2, *Tv*TPS3, and *Tv*TPS9 were created with Gateway® technology (Invitrogen). Sequences were subcloned from the pCR4-TOPO vector into the pDONR207 vector with BP Clonase II. Subsequently, the constructs were cloned into the bacterial expression vector pH9GW, which contains a T7-promoter and a N-terminal 9 x His-tag, with LR Clonase II according to the manufacturer's instructions. The expression constructs were then verified by sequencing and transformed into the *E. coli* BL21(DE3) expression strain, which contains a copy of the T7-RNA-polymerase.

For the expression of recombinant protein, an overnight culture of the desired clone was inoculated 1:20 in 100 ml LB-medium supplemented with 50 µg ml^{-1} kanamycine and was allowed to grow to an OD_{600} of 0.5 – 0.7 at 37 °C. The vector mediated expression was induced by adding IPTG (isopropyl-*β*-thiogalactopyranose) to a final concentration of 1 mM and the culture was incubated over night at 18 °C, to increase plasmid stability, allow best possible protein folding, and reduce the possibility of inclusion body formation (Jonasson et al. 2002).

The cells were harvested by centrifugation at 4 °C for 5 min at 5000 x g, washed with 5 ml ice-cold extraction buffer (50 mM MOPSO pH 7.0, 5 mM $MgCl_2$, 5 mM DTT, 5 mM sodium ascorbate, 0.5 mM PMSF (phenylmethylsulphonyl fluoride), 10 % [v/v] glycerol), and resuspended in 3 ml ice-cold extraction buffer. The cellsuspension was then disrupted on ice at 60 % power for 4 x 30 s with a sonicator (Bandelin UW2070, Berlin, Germany), with 30 s pauses between pulses. The cell fragments were then removed by centrifugation at 4 °C

for 30 min at 21.000 x g, and a buffer exchange into reaction buffer (10 mM MOPSO pH 7.0, 1 mM DTT, 10 % [v/v] glycerol) was carried out by passing the enzyme containing solution over an equilibrated Econo-Pac® 10 DG column (Bio-Rad Laboratories, Hercules, USA).

Sequencing and bioinformatic analysis

Sequencing of DNA fragments was performed on an ABI3100 sequencer (Applied Biosystems, Foster City, USA) using the method described by Sanger et al. (1977). DNA and protein sequences were processed with the SeqMan program (Lasergene DNAStar V8.02, Madison, USA), ClustalX2 (Larkin et al. 2007), and BioEdit 7.0.9. (Hall 1999). Sequence databank searches and sequence analysis were performed using BLAST via the NCBI sequence database (http://blast.ncbi.nlm.nih.gov/Blast.cgi), signal peptides were predicted with the programs TargetP 1.1 (Emanuelsson et al. 2007), ChloroP 1.1 (Emanuelsson et al. 1999), SignalP 3.0 (Emanuelsson et al. 2007, Petersen et al. 2011) (http://www.cbs.dtu.dk/services/), as well as the program Predotar (Small et al. 2004) (https://urgi.versailles.inra.fr/Tools/Predotar), while transmembrane helices were predicted with TMHMM 2.0 (Krogh et al. 2001) (http://www.cbs.dtu.dk/services/).

The phylogenetic trees were constructed by the Neighbor-Joining method (10.000 bootstrap replicates) using the ClustalX2 program (protein weight matrix: Gonnet series, gap penalty: 10.00, gap length penalty: 0.20, delay divergent sequences: 30 %) and the TreeView 1.6.6 software (Page 1996), and adjusted with Adobe Illustrator CS5 (Adobe Systems Inc., San Jose, USA).

Models of the three-dimensional structure of *Tv*TPS1 – *Tv*TPS7 were generated using the SWISS-MODEL server (Arnold et al. 2006). For modelling, the amino acid sequences were fitted to the template structure of *Salvia officinalis* bornyl diphosphate synthase (Whittington et al. 2002). The resulting models were visualized with the program PyMOL 0.99 UCSF Chimera beta version 1 (http://www.pymol.org/).

Results

Isolation and sequence comparison of cDNA clones encoding putative terpene synthases of thyme.

In order to identify the terpene synthases responsible for monoterpene chemotype formation in common garden thyme, we generated cDNA libraries from the chemotypes T_{28} and A_{10}, whose main monoterpenes are thymol and carvacrol and their presumed precursor γ-terpinene in T_{28}, and α-terpineol in A_{10}. The glandular trichome cell clusters, which are the site of monoterpene biosynthesis in the Lamiaceae, were sheared off the leaf surfaces, purified and used for RNA extraction to ensure a high concentration of cDNA sequences related to monoterpene biosynthesis. These two libraries were then used for colony screening with [α^{32}P]dATP labeled probes. In addition to this, RACE cDNA-libraries for all chemotypes (G_1, A_{10}, U_6, L_{48}, T_2, C_{10}, and T_{28}) were generated and searched with the degenerate primers Lab1-fwd (5'-CA[A/G] TTG GAG TTG ATT GAT GAC TTG-3'), Lab2-fwd (5'-CTC TTG GAT TCA GAC T[C/T]C TCA GAC AAC ATG GTT T-3'), and Lab3-fwd (5'-GTG AGG GAT AGG CT[A/G/C/T] GTG GAA [A/T]GC TAC TTT TGG-3'). These primers were constructed based on conserved regions found in known monoterpene synthase genes from different Lamiaceae species. The EST (expressed sequence tag) fragments isolated from colony screening and from the utilization of degenerate primers were extended by several rounds of RACE-PCR to yield four individual full length cDNA clones which were designated as the putative terpene synthases *Tvtps1*, *Tvtps2*, *Tvtps3*, and *Tvtps9*. Four individual cDNA clones which could not be extended to full length by several rounds of RACE-PCR were designated as the putative terpene synthases *Tvtps5* trough *Tvtps8*. With all cDNA, RACE-PCR was repeated at least twice to confirm the 5'-end of the open reading frame, when available.

Monoterpene synthases isolated from *T. vulgaris*.

The isolated cDNA clones of the putative monoterpene synthases *Tvtps1*, *Tvtps2*, and *Tvtps3* had lengths of 2102 bp, 2133 bp, and 2181 bp, respectively, with open reading frames (ORF) of 1791 bp, 1788 bp, and 1740 bp. These encode proteins of 597 aa, 596 aa, and 580 aa length. The estimated molecular weight and calculated isoelectric points for the proteins are 68.9 kDa and pI 5.0 for *Tvtps1*, 68.9 kDa and pI 5.7 for *Tvtps2*, and 67.4 kDa and pI 5.3 for *Tvtps3*. The sequences of the putative monoterpene synthases *Tvtps5*, *Tvtps6*, and *Tvtps7* could only be partially isolated and were missing parts of the N-terminal region. The full

length sequence of *Tvtps5* was later successfully isolated by Jette Schimmel (Martin-Luther University of Halle-Wittenberg, Germany) and encodes for a 599 aa long protein, with an estimated molecular weight of 69.5 kDa and a calculated isoelectric point of pI 5.2 (personal communication). Sandra Krause (Martin-Luther University of Halle-Wittenberg, Germany) successfully isolated enough of the missing portions of the transit peptides of *Tvtps6* (560 aa) and *Tvtps7* (554 aa), including the double RR, to achieve successful heterologous expression (Krause et al. 2013). The sequences of *Tvtps1* (46 aa), *Tvtps2* (54 aa), and *Tvtps5* (49 aa) each contain a putative transit peptide for plastid targeting, but this was not true for *Tvtps3* and no prediction could be made for the abridged sequences of *Tvtps6* and *Tvtps7*, using the prediction programs TargetP 1.1 and ChloroP 1.1. Lacking a transit peptide, *Tvtps3* is likely to be targeted to the cytosol rather than the plastid.

All isolated sequences contained several sequence motifs characteristic for monoterpene synthases (Figure 3-1 and Figure 3-2). For example, the double arginine motif (**RRx$_8$W**), commonly found in the N-terminal domain of terpene cyclases, is situated at residues 56 - 66 in *Tvtps1*, at residues 57 - 67 in *Tvtps2*, and at residues 38 - 48 in *Tvtps3*, respectively. In *Tvtps5*, *Tvtps6*, and *Tvtps7* the motif is present at 58 - 68 aa, at 18 - 28 aa, and at 18 - 28 aa, respectively. The **RxRx$_6$W motif**, situated at position 312 - 321 in *Tvtps1*, 315 - 324 in *Tvtps2*, and 299 – 308 in *Tvtps3*, can be found at position 315 - 324 in *Tvtps5*, 277 - 286 in *Tvtps6*, and 273 - 282 in *Tvtps7*, respectively. This motif is thought to be involved in the complexation of the diphosphate function after ionization of the substrate thus preventing premature nucleophilic attack on the carbocationic intermediates (Starks et al. 1997). Situated in the C-terminal domain of terpene synthases, 27 aa downstream of the RxRx$_6$W motif, is the aspartate-rich **DDxxD motif**, a highly conserved motif involved in binding the bivalent metal ion cofactor (Aaron and Christianson 2010). This important motif was found to be present in all of the genes isolated. It was situated at residues 349 – 353 in *Tvtps1*, 352 – 356 in *Tvtps2*, 336 – 340 in *Tvtps3*, and at 352 – 356, 314 – 318, and 310 – 314 in *Tvtps5*, *Tvtps6*, and *Tvtps7*, respectively. The second aspartate-rich motif, the **NSE/DTE motif**, also involved in the binding of the metal cofactor (Aaron and Christianson 2010), is present in all thyme sequences as D-D-x-G-T-x-x-x-E at residues 493 - 501 in *Tvtps1*, 495 - 503 in *Tvtps2*, and 479 - 487 in *Tvtps3*, while it can be found at position 496 - 504 in *Tvtps5*, 458 - 466 in *Tvtps6*, and 452 – 460 in *Tvtps7*, respectively.

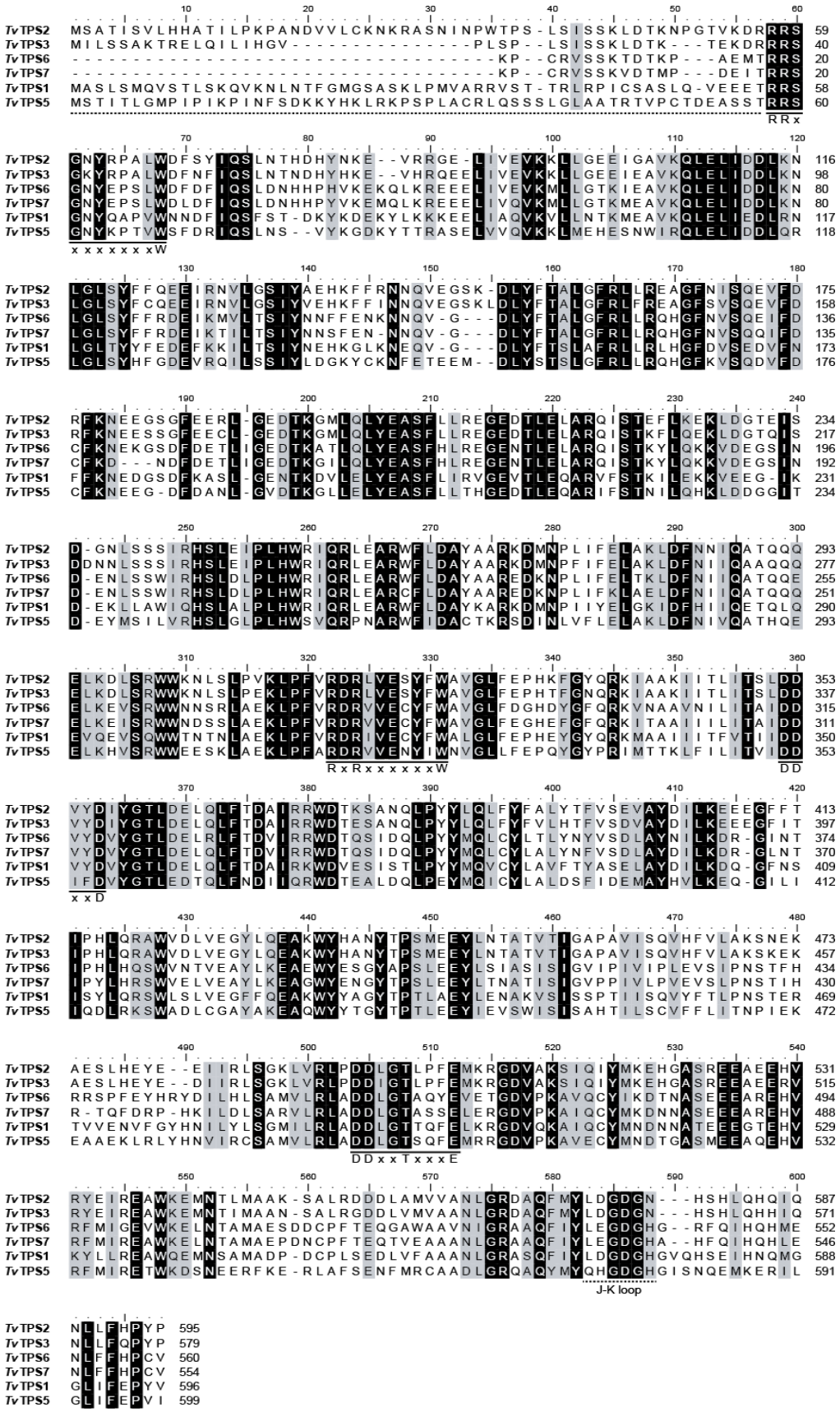

TvTPS2
TvTPS3
TvTPS6
TvTPS7
TvTPS1
TvTPS5
RRx
xxxxxxxW
RxRxxxxxxW
DD
xxD
DDxxTxxxE
J-K loop

Figure 3-1: Alignment of the deduced amino acid sequences of *Thymus vulgaris* monoterpene synthase genes. Included are *Tv*TPS1 (*γ*-terpinene synthase), *Tv*TPS2 (linalool synthase), *Tv*TPS3 (linalool synthase), *Tv*TPS5 (putative monoterpene synthase), *Tv*TPS6 (*cis*/*trans*-sabinene hydrate synthase), and *Tv*TPS7 (*trans*-sabinene hydrate synthase). Amino acids which are common to all sequences are shaded black, while residues with similar properties are shaded in grey. The RRx_8W motif, the $RxRx_6W$ motif, as well as the metal binding motifs DDxxD and NSE/DTE are indicated by solid bars. The J-K loop is indicated by a dotted bar. The full length sequences of *Tv*TPS6 and *Tv*TPS7 were isolated by Sandra Krause (2013), and the full length sequence of *Tv*TPS5 was isolated by Jette Schimmel (personal communication). The alignment was created with the ClustalX2 program using the default settings, adjusted by hand in the BioEdit program, and edited in Adobe Illustrator CS.

All six putative monoterpene synthase sequences displayed high similarity to known plant monoterpene synthase sequences based on comparison at the amino acid level (Table 3-1). The highest similarity of *Tvtps1* was to the sequences of the *Origanum vulgare γ*-terpinene synthase (91.5 % identity) and to the *Salvia pomifera* sabinene synthase (67.2 % identity). *Tvtps2* and *Tvtps3*, on the other hand, showed the highest identity to the linalool synthases of *Mentha citrata* (78.2 and 77.9 % identity, respectively), *Perilla citriodora* (70.5 and 70.6 % identity, respectively), and *P. frutescens* (70.9 and 70.6 % identity, respectively). Both putative monoterpene synthases, *Tvtps6* and *Tvtps7*, displayed the highest similarity to the *O. vulgare* sabinene synthase (87.8 and 86.4 % identity, respectively) and the *O. vulgare* (*E*)-*β*-ocimene synthase (82.6 and 82.5 % identity, respectively). In contrast to this, the sequence of *Tvtps5* showed comparatively low levels of identity (max. 55.9 %) to other known plant monoterpene synthases with the exception of the *Rosmarinus officinalis* limonene synthase, to which it had an identity of 76.0 %. When the predicted target peptides were truncated before the double arginine of the RRx_8W motif, the identity between the putative *T. vulgaris* monoterpene synthase sequences varied between 49.7 and 92.2 %. The *Tvtps5* again showed comparatively low identity levels (approx. 50 %) to the other putative monoterpene synthases from *T. vulgaris*, while the identity percentages of *Tvtps1* to the other sequences ranged from 58 – 64 %. Inside the clusters of *Tvtps2* and *Tvtps3* as well as *Tvtps6* and *Tvtps7* the relative identities were quite high: 92.2 and 85.8 %, respectively (Table 3-1).

Organism		Identity on the amino acid level [%]						Main synthase product
		TPS1	**TPS2**	**TPS3**	**TPS5**	**TPS6**	**TPS7**	
T. vul.	***Tv*TPS1**	100						*γ*-terpinene
T. vul.	***Tv*TPS2**	58.5	100					linalool
T. vul.	***Tv*TPS3**	58.2	92.2	100				linalool
T. vul.	***Tv*TPS5**	51.4	50.2	49.7	100			inactive
T. vul.	***Tv*TPS6**	60.3	61.9	61.6	52.8	100		sabinene hydrate
T. vul.	***Tv*TPS7**	64.0	64.5	64.4	53.7	85.8	100	sabinene hydrate
O. vul.	GU385978	**91.5**	60.5	59.5	51.8	62.5	65.9	*γ*-terpinene
S. pom.	DQ785794	67.2	63.2	62.6	55.3	64.7	68.1	sabinene
R. off.	ABP01684	66.6	65.5	65.0	55.9	66.5	69.7	pinene
R. off.	AF051901	66.3	62.5	61.7	55.3	64.2	68.3	sabinene
P. fru.	FJ644547	62.4	65.1	64.9	53.7	59.6	61.6	geraniol
P. cit.	AY917193	62.2	**70.5**	**70.6**	54.4	62.8	67.0	linalool
P. fru.	AF444798	61.8	**70.9**	**70.6**	54.1	62.6	66.0	linalool
O. vul.	GU385980	61.5	62.5	61.4	53.8	**87.8**	**86.4**	sabinene
M. cit.	AY083653	60.9	**78.2**	**77.9**	53.3	62.7	65.8	linalool
S. off.	AF051900	60.9	62.7	62.8	52.7	60.6	63.3	(+)-bornyl diphosphate
L. ang.	ABB73044	60.9	61.0	60.4	53.4	60.8	62.8	limonene
L. lat.	ABD77417	60.7	60.6	60.1	53.4	60.8	62.8	linalool
O. bas.	AAV63791	60.6	61.8	61.6	51.8	59.6	62.5	myrcene
O. vul.	GU385967	60.0	60.8	60.0	51.7	**82.6**	**82.5**	(*E*)-*β*-ocimene
O. bas.	AAV63790	60.0	60.6	60.8	50.7	58.4	61.2	fenchol
L. ang.	ABB73045	58.3	67.0	66.9	53.0	57.7	60.6	linalool
O. bas.	AAV63792	57.4	59.3	58.7	49.2	59.9	61.9	terpinolene
O. fra.	ACM92062	52.3	59.7	59.1	50.7	54.1	57.2	linalool
S. ste.	AF527416	50.8	51.4	50.0	54.3	51.3	53.2	3-carene
R. off.	ABI20515	49.9	50.6	51.0	52.9	48.7	52.1	cineol
P. cit.	AF233894	49.7	51.3	51.0	54.1	52.4	54.2	limonene
S. off.	AF051899	49.6	50.8	50.3	51.8	49.6	52.3	1,8-cineole
S. fru.	FJ618810	49.5	49.3	49.1	53.4	52.4	52.5	1,8-cineole
R. off.	ABD77416	49.2	51.7	51.6	**60.0**	51.5	52.3	limonene
C. uns.	BAD27258	42.6	45.4	43.7	47.7	45.3	46.0	*γ*-terpinene
C. lim.	AAM53943	42.4	45.0	43.7	47.2	44.5	45.2	*γ*-terpinene

Table 3-1: Sequence identity levels of monoterpene synthases from *T. vulgaris* with each other and functionally related monoterpene synthases. GenBank accession numbers are given after the species name. Abbreviations are as follows: *Tv*TPS1 (γ-terpinene synthase), *Tv*TPS2 (linalool synthase), *Tv*TPS3 (linalool synthase), *Tv*TPS5 (putative monoterpene synthase), *Tv*TPS6 (*cis*/*trans*-sabinene hydrate synthase), and *Tv*TPS7 (*trans*-sabinene hydrate synthase), *C. lim.* - *Citrus limon*, *C. uns.* - *Citrus unshiu*, *M. cit.* - *Mentha citrata*, *L. ang.* - *Lavandula angustifolia*, *L. lat.* - *Lavandula latifolia*, *O. bas.* - *Ocimum basilicum*, *O. vul.* - *Origanum vulgare*, *O. fra.* - *Osmanthus fragrans*, *P. cit.* - *Perilla citriodora*, *P. fru.* - *Perilla frutescens*, *R. off.* - *Rosmarinus officinalis*, *S. fru.* - *Salvia fruticosa*, *S. off.* - *Salvia officinalis*, *S. ste.* - *Salvia stenophylla*, *T. vul.* - *Thymus vulgaris*.

Functionally important regions for the biosynthesis of monoterpenes

To identify the functionally important regions of monoterpene synthase genes, domain swapping (El Tamer et al. 2003, Peters and Croteau 2003), site-directed mutagenesis (Starks et al. 1997, Degenhardt et al. 2009), and structural modeling-based approaches (Kampranis et al. 2007) have been applied in several studies. To examine the active-site region of the monoterpene synthases isolated from thyme, the predicted secondary and tertiary structures were compared with the *S. officinalis* bornyl pyrophosphate synthase crystal structure (Whittington et al. 2002), which revealed a number of residues that were highly conserved and others that varied and therefore could play an important role in product specificity of the synthases (Figure 3-2). The enzyme regions examined include the tyrosine residue (Y) found in the RRx_8W motif, the $RxRx_6W$ motif, the DDxxD motif, the NSE/DTE motif, the J-K loop, two regions with clustered variable residues (termed Region 1 and Region 2) (Kampranis et al. 2007), as well as a third region (termed Region 3) not belonging to a known motif or previously named region.

The tyrosine, found in the RRx_8W motif situated in the N-terminal region of the enzyme, is present in all examined enzyme sequences, and the orientation of the residue appears to be similar in all the examined enzymes. The $RxRx_6W$ motif in turn, situated partially on the A-C loop and the helix-C, reaches deep into the active center with the first arginine (R) and tryptophan (W) residues. The DDxxD motif is composed of identical amino acids, with the exceptions of the inactive *Tv*TPS5, the *R. officinalis* limonene synthase, and the *S. officinalis* 1,8-cineole synthase. These three synthases possess a phenylalanine (F) instead of a tyrosine (Y), and *Tv*TPS5 additionally possesses an isoleucine (I) instead of a valine (V). The orientation of the aspartate (D) residues, however, appears to be similar in the examined sequences. Closer examination of the NSE/DTE motif, situated on helix-H, shows

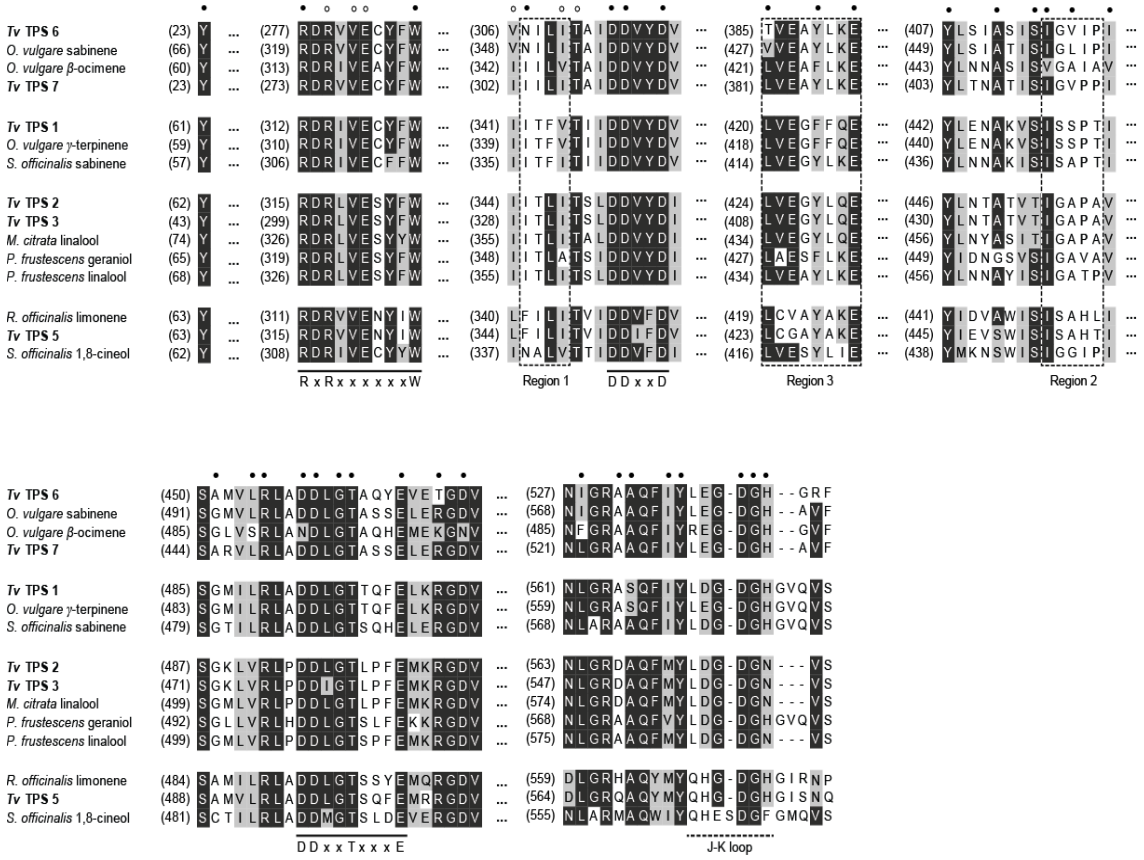


Figure 3-2: Multiple sequence alignment of deduced amino acid regions involved in the active center of Lamiaceae monoterpene synthases. The enzymes included are thyme *Tv*TPS1 (*γ*-terpinene synthase), *Tv*TPS2 (linalool synthase), *Tv*TPS3 (linalool synthase), *Tv*TPS5 (inactive protein), *Tv*TPS6 ((*Z*)-sabinene hydrate synthase), and *Tv*TPS7 ((*E*)-sabinene hydrate synthase) with their nearest homologues and functionally similar monoterpene synthases. Amino acids which are common to thirteen or more sequences are shaded in black, while residues with similar properties are shaded in grey. The $RxRx_6W$ motif, the DDxxD motif, as well as the NSE/DTE motif are indicated by solid bars. The J-K loop is indicated by a dotted bar, while Region 1, Region 2, and Region 3 are highlighted by a dotted frame. Solid black dots mark amino acids whose R-groups are predicted to extend into the active site, while black circles mark amino acids predicted to be in close proximity to the active site. The alignment was created with ClustalX2 using the default settings, adjusted by hand in BioEdit, and edited in Adobe Illustrator CS.

that with the exception of the *O. vulgare* (*E*)-*β*-ocimene synthase (Asn-493 instead of Asp-493), all residues of this motif thought to be involved in enzyme activity are the same. However, the amino acid residues in the vicinity of this motif are quite diverse, and no pattern was observed that is correlated to the activity of the enzymes. Region 1 and 2 are not

part of known motifs but have been shown to influence the enzyme activity of monoterpene synthases (Kampranis et al. 2007). Region 1 is situated on helix-D at the bottom of the active site cavity, about two amino acids upstream of the DDxxD motif, and comprises four residues (Figure 3-2). The first residue in Region 1 is an asparagine (A) in the sequences of *Tv*TPS6 (Asn-307), *O. vulgare* sabinene synthase (Asn-349), and *S. officinalis* 1,8-cineole synthase (Asn-338), while the corresponding residues in the other thyme enzymes are an isoleucine (I) in the case of *Tv*TPS1 (Ile-342), *Tv*TPS2 (Ile-345), *Tv*TPS3 (Ile-329), and *Tv*TPS7 (Ile-303) and a phenylalanine (F) in the case of *Tv*TPS5 (Phe-345) and *R. officinalis* limonene synthase (Phe-345), respectively. The conservation of an asparagine in the corresponding position of the 1,8-cineole synthases of *S. officinalis*, *S. fruticosa*, and *Arabidopsis thaliana* suggests that it may be a critical residue in the catalytic mechanism for the biosynthesis of this monoterpene (Chen et al. 2004, Kampranis et al. 2007, Roeder et al. 2007). The side chain of this residue is thought to be hydrogen bonded to a water molecule facilitating the attack on the α-terpinyl cation which results in the formation of α-terpineol (Figure 3-11)(Kampranis et al. 2007). Region 2 comprises five residues, which are part of the loop connecting helices G1 and G2. The presence of a proline (P) at position 450 in the 1,8-cineole synthase of *S. officinalis* has been shown to disrupt the hydrogen bonding network and thereby induce the formation of a 'kink' between the helices G1 and G2, thereby exposing the carbonyl oxygens of Ile-446 and Gly-447 to the active-site cavity (Kampranis et al. 2007). A proline in the corresponding position can also be found in *Tv*TPS6 (Pro-419), *Tv*TPS7 (Pro-415), as well as in the *O. vulgare* sabinene synthase (Pro-461) and the *P. frutescens* linalool synthase (Pro-468). It appears to have the same disrupting influence on the hydrogen bonding network as described for the *S. officinalis* 1,8-cineole synthase. Since it is thought that stabilization of the highly reactive carbocationic intermediates by local partial charges in the catalytic pocket is an essential part of the catalytic mechanisms of terpene biosynthesis, it is likely that 'kink' is an essential structural feature of these enzymes. Similar deformations have been previously reported for enzymes as different as the *Fusarium sporotrichoides* trichodiene synthase, the *Penicillium roqueforti* aristolochene synthase, and the already mentioned *S. officinalis* 1,8-cineole synthase (Kampranis et al. 2007).

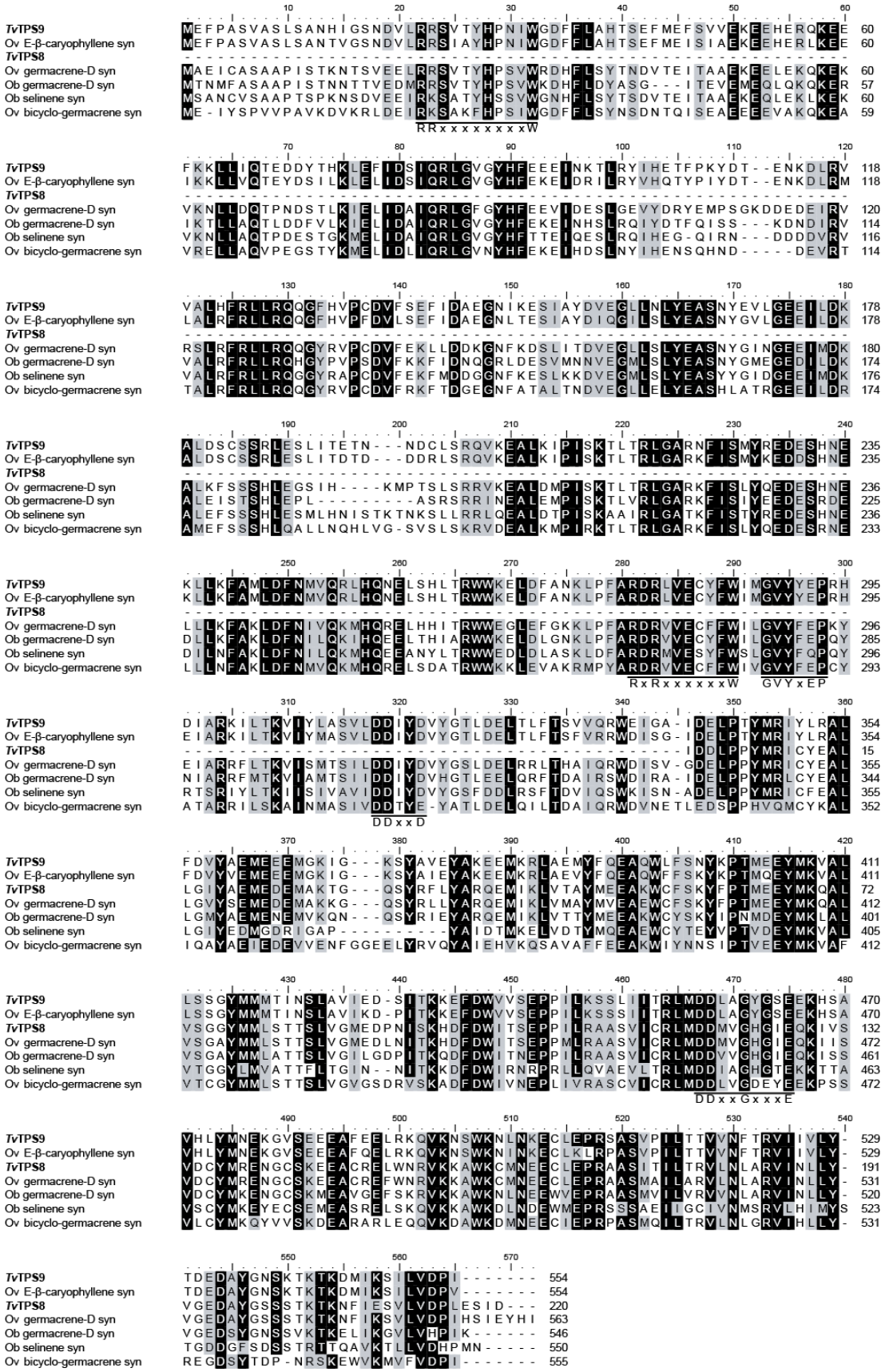
Three nonconsecutive amino acids, found on helix-F, also form part of the active center but are not part of or near a known motif. With the exception of *Tv*TPS6 (Thr-385) and the *O. vulgare* sabinene synthase (Val-427), the corresponding position of the first amino acid in all other synthases is occupied by leucine (L). The second residue, which is in close

proximity to the substrate, is in all cases an aromatic amino acid. While phenylalanine (F) can be found in the corresponding position in the sequences of the *Tv*TPS1 (Phe-424), the *O. vulgare* (*E*)-*β*-ocimene synthase (Phe-425), the *O. vulgare* *γ*-terpinene synthase (Phe-422), and the *P. frutescens* geraniol synthase (Phe-431), this residue is a tyrosine (Y) in all other sequences. The third amino acid in all examined sequences is glutamic acid (E) (Region 3 in Figure 3-2). The positioning of the J-K loop has been shown to play an important role in the enzyme activity of monoterpene synthases (Crowell et al. 2002, van Schie et al. 2007). Deletion of single amino acids can have significant influence on the formation of the loop and thus on the accessibility of the active site pocket.

Sesquiterpene synthases isolated from *T. vulgaris*

The cDNA clone of *Tvtps9* was first isolated with primers originally designed for the (*E*)-*β*-caryophyllene synthase from *O. vulgaris* (Crocoll 2011). The isolated sequence codes for a deduced amino acid sequence of 554 aa length, with an estimated molecular weight of 64.8 kDa and a calculated isoelectric point of pI 5.0. Consistent with the assumed cytosolic location of sesquiterpene biosynthesis, no N-terminal organelle targeting information could be found. The sequence of *Tvtps8*, a putative sesquiterpene synthase, however could not be isolated in full length to date. The amino acid sequence that can be deduced from the partial nucleotide sequence is 217 aa long and ends about 50 aa downstream from the putative location of the DDxxD motif. The partial sequence of *Tvtps8* could be isolated with primers originally designed for the germacrene D synthase from *O. vulgaris* (Crocoll 2011).

Figure 3-3: Alignment of the deduced amino acid sequences of *T. vulgaris* sesquiterpene synthase genes. Included are the thyme *TvTPS8* (partial sequence of a putative germacrene D synthase) and *TvTPS9* ((*E*)-*β*-caryophyllene synthase) predicted proteins with their nearest homologues and functionally similar sesquiterpene synthases. Ov (*E*)-*β*-caryophyllene synthase (GU385970), Ov germacrene D synthase (GU385976), and Ov bi-cyclo-germacrene synthase (GU385973) are from *Origanum vulgare*, while Ob germacrene D synthase (AAV63786) and Ob selinene synthase (AAV63785) are sesquiterpene synthases from *Ocimum basilicum*. Amino acids which are common to six or more sequences are shaded in black, while residues with similar properties are shaded in grey. The RRx_8W motif, the $RxRx_6W$ motif, as well as the metal binding motifs DDxxD and NSE/DTE, which are highly conserved in plant terpene synthases are indicated by solid bars, as is the GVYxEP element common to angiosperm sesquiterpene synthases. The alignment was created with ClustalX2 using the default settings, adjusted by hand in BioEdit, and edited in Adobe Illustrator CS.


TvTPS9
Ov E-β-caryophyllene syn
TvTPS8
Ov germacrene-D syn
Ob germacrene-D syn
Ob selinene syn
Ov bicyclo-germacrene syn
RRxxxxxxxxW
RxRxxxxxxW
GVYxEP
DDxxD
DDxxGxxxE

The deduced amino acid sequences of *Tvtps9* and *Tvtps8* both contain, like the sequences of *Tvtps1* through *Tvtps7*, characteristic terpene synthase motifs (Figure 3-3). The RRx_8W motif, the $RxRx_6W$ motif, the **GVYxEP element**, commonly found in angiosperm sesquiterpene synthases (Cai et al. 2002), and the previously mentioned metal-binding DDxxD motif are all present in the deduced amino acid sequence of *Tvtps9* at residues 22 - 32, 276 - 284, 288 - 293, and 313 - 317, respectively. The NSE/DTE motif, the second aspartate-rich motif involved in binding the metal cofactor, is present in *Tvtps8* and *Tvtps9* at residues 119 - 127 and 457 - 465, respectively. This motif is present in both sequences as D-D-x-x-G-x-x-x-E, the glycine replacing the more common serine (S) or threonine (T), thus facilitating the coordination of an additional water molecule to the metal cofactor (Aaron and Christianson 2010).

Only a few sesquiterpene synthases have been described to date for the plant family Lamiaceae, and the deduced amino acid sequences of *Tvtps8* and *Tvtps9* showed the highest similarities to them (Table 3-2). The partial sequence of *Tvtps8* displayed the highest identity to the germacrene D synthases of *O. vulgare* (88.0 % identity) and *Ocimum basilicum* (74.7 % identity) while it only shared identities of 50.2 % and 54.8 % with the *O. basilicum* selinene synthase and the *O. vulgare* (*E*)-*β*-caryophyllene synthase, respectively. *Tvtps9* in turn displayed the highest identity (88.4 %) to the *O. vulgare* (*E*)-*β*-caryophyllene synthase, while sharing comparatively low identities (51.5 – 58.3 %) with the remaining Lamiaceae sesquiterpene synthases, as well as with the intraspecific *Tvtps8* sequence (55.9 % identity).

Organism		Identity on the amino acid level [%]		Main synthase product
		***Tv*TPS8**	***Tv*TPS9**	
T. vulgaris	***Tv*TPS8**	100		*n.d.*
T. vulgaris	***Tv*TPS9**	55.9	100	(*E*)-*β*-caryophyllene
O. vulgare	GU385970	54.8	**88.4**	(*E*)-*β*-caryophyllene
L. angustifolia	AGL98419	56.0	**68.7**	(*E*)-*β*-caryophyllene
O. vulgare	GU385976	**88.0**	58.3	germacrene D
O. basilicum	AY693644	**74.7**	56.8	germacrene D
O. basilicum	AY693643	50.2	51.5	selinene
L. angustifolia	AGL98420	56.5	51.2	germacrene D
V. vinifera	AY561842	52.5	50.0	germacrene D
V. vinifera	JF808010	53.0	49.8	(*E*)-*β*-caryophyllene
C. creticus	EU726270	43.8	48.1	germacrene B
P.tri. x P.del.	AAR99061	46.1	48.1	germacrene D
A. deliciosa	AAX16121	49.5	47.4	germacrene D
C. hystrix	ADX01384	42.2	47.3	germacrene D
S. lycopersicum	ADD96698	47.1	47.2	*β*-caryophyllene/*α*-humulene

Table 3-2: Sequence identity levels of sesquiterpene synthases from *Thymus vulgaris* with each other and functionally related sesquiterpene synthases. *A. deliciosa* - *Actinidia deliciosa* (Actinidiaceae), *C. creticus* - *Cistus creticus* (Cistaceae), *C. hystrix* - *Citrus hystrix* (Rutaceae), *L. angustifolia* – *Lavandula angustifolia* (Lamiaceae), *O. basilicum* - *Ocimum basilicum* (Lamiaceae), *O. vulgare* - *Origanum vulgare* (Lamiaceae), *P.tri*. x *P.del.* – *Populus trichocarpa* x *P. deltoides* (Salicaceae), *S. lycopersicum* – *Solanum lycopersicum* (Solanaceae), *T. vulgaris* - *Thymus vulgaris* (Lamiaceae), *V. vinifera* - *Vitis vinifera* (Vitaceae).

Terpene synthases isolated from *T. vulgaris* produce a variety of mono- and sesquiterpenes

The putative terpene synthase genes were expressed in a bacterial system to confirm and characterize their enzymatic activity. A C-terminal His-tag was added to facilitate enzyme purification from the bacterial extract. Previous studies on monoterpene synthases from members of the Lamiaceae have shown that the removal of the sequence encoding the N-terminal region upstream of the RRx_8W motif is sometimes required in order to obtain an active enzyme after heterologous expression in *E. coli* (Williams et al. 1998, Crocoll et al. 2010). Taking these considerations into account, the putative monoterpene synthases *Tv*TPS1, *Tv*TPS2, and *Tv*TPS3 were expressed in truncated and non-truncated forms. *Tv*TPS1 was active both as a full length protein and after the truncation of the 55 amino acids upstream of the double arginine motif. *Tv*TPS1 converted geranyl diphosphate (GPP) to the monoterpene olefin *γ*-terpinene (89.6 % of total products) with minor amounts of *α*-terpinene (6.1 %), *α*-thujene (1.5 %), myrcene (1.0 %), limonene (0.7 %), *α*-phellandrene (0.4 %), *α*-pinene (0.3 %), and sabinene (0.3 %), as well as trace amounts of *p*-cymene (Figure 3-4 A). Both putative monoterpene synthases *Tv*TPS2 and *Tv*TPS3 converted GPP to linalool. However, while *Tv*TPS2 was only active in its truncated form, the same product profile could be observed for the full-length protein of *Tv*TPS3 and after the N-terminal deletion of 56 amino acids (B and C of Figure 3-4). Farnesyl diphosphate (FPP), the precursor of sesquiterpenes, was not converted into sesquiterpene products by *Tv*TPS1, *Tv*TPS2, nor *Tv*TPS3 (data not shown).

The sequences of *Tvtps6* and *Tvtps7*, which could not be isolated in full length but ended just upstream of the RRx_8W motif, were successfully expressed in truncated form by Sandra Krause (2013). The enzymes *Tv*TPS6 is a multiproduct enzyme that converted GPP to (*Z*)-sabinene hydrate (60 % of total products) and (*E*)-sabinene hydrate (21 %) as well as the minor products *α*-pinene (3.2 %), myrcene (2.6 %), limonene (2 %), *α*-thujene (1.8 %), *γ*-terpinene (1.8 %), *α*-terpineol (1.7 %), *β*-pinene (1.3 %), sabinene (1.1 %), *α*-terpinene (0.8

%), and terpinene-4-ol (0.8 %), and trace amounts of terpinolene, *α*-phellandrene, camphene, and (*E*)-*β*-ocimene (Table 3-3). An almost identical product profile was observed for *Tv*TPS7. While differing strongly in the distribution of (*E*)- and (*Z*)-sabinene hydrate (78 to 1 % of total products, respectively), the distribution of the minor products limonene (6.3 %), *α*-terpineol (3.8 %), *α*-pinene (3.6 %), *γ*-terpinene (1.4 %), myrcene (1.2 %), sabinene (1 %), *α*-terpinene (0.7 %), α-thujene (0.6 %), *β*-pinene (0.6 %), and terpinene-4-ol (0.6 %), as well as trace amounts of terpinolene, *α*-phellandrene, and camphene, differed only slightly from that observed for *Tv*TPS6 (Table 3-3).

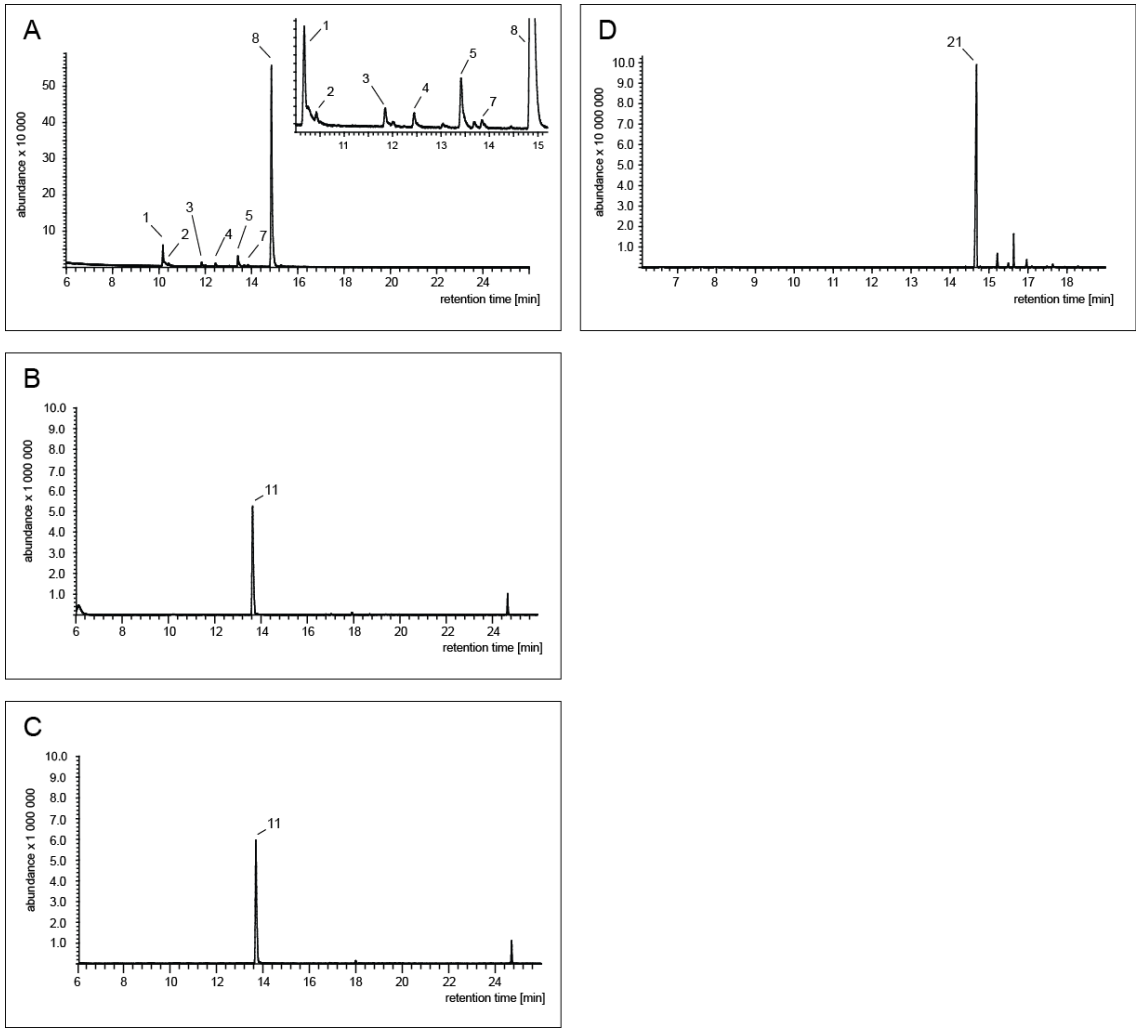


Figure 3-4: Terpene products of *Tv*TPS1, *Tv*TPS2, *Tv*TPS3, and *Tv*TPS9 measured *in vitro*. The enzymes were expressed in *E. coli*, extracted, and incubated with the substrate GPP (FPP in the case of *Tv*TPS9) and 20 mM Mg^{2+}ions. The resulting terpene products were identified by gas chromatography coupled to mass spectrometry; the total ion chromatogram is shown. (**A**) Products of *Tv*TPS1 with a

truncation of the N-terminal 55 amino acids: 1. *α*-thujene, 2. *α*-pinene, 3. sabinene, 4. myrcene, 5. *α*-terpinene, 7. limonene, 8. *γ*-terpinene. (**B**) Products of *Tv*TPS2 with a truncation of the N-terminal 57 amino acids: 11. linalool. (**C**) Products of *Tv*TPS3 with a truncation of the N-terminal 37 amino acids: 11. linalool. (**D**) Products of *Tv*TPS9: 21. (*E*)-*β*-caryophyllene.

Monoterpene	*Tv*TPS6 [%]	*Tv*TPS7 [%]
α-thujene	1.8	0.6
α-pinene	3.2	3.6
camphene	0.3	0.2
sabinene	1.1	1.0
β-pinene	1.3	0.6
myrcene	2.6	1.2
α-phellandrene	0.4	0.3
α-terpinene	0.8	0.7
limonene	2.0	6.3
(*E*)-*β*-ocimene	0.4	n.e.i.
c-terpinene	1.8	1.4
(*E*)-sabinene hydrate	20.9	77.8
terpinolene	0.5	0.4
(*Z*)-sabinene hydrate	60.4	1.4
terpinene-4-ol	0.8	0.6
α-terpineol	1.7	3.8

Table 3-3: Monoterpene composition of the product blends formed by *Tv*TPS6 and *Tv*TPS7. The amounts are given as a percentage of the total monoterpene content. N.e.i. = not elsewhere identified. Table was adapted from Krause et al. (2013).

The putative terpene synthase gene *Tvtps9*, considered to be a sesquiterpene synthase as it shared a high sequence identity with known sesquiterpene synthases, was shown to convert FPP to (*E*)-*β*-caryophyllene (Figure 3-4 D).

Phylogeny of terpene synthases of *T. vulgaris*

To obtain further information about the evolutionary relationships of the cloned terpene synthases, the predicted amino acid sequences were analyzed and assembled into a phylogenetic tree. To this end, biochemically characterized terpene synthases of four of the seven defined TPS classes as well as the sequences isolated from *T. vulgaris*, were analyzed with ClustalX2. The analyzed sequences also comprised sequences showing the highest homology in a BLAST search to the sequences isolated in this study. Using the Neighbor-Joining method, a phylogenetic tree was calculated and constructed using ClustalX2.

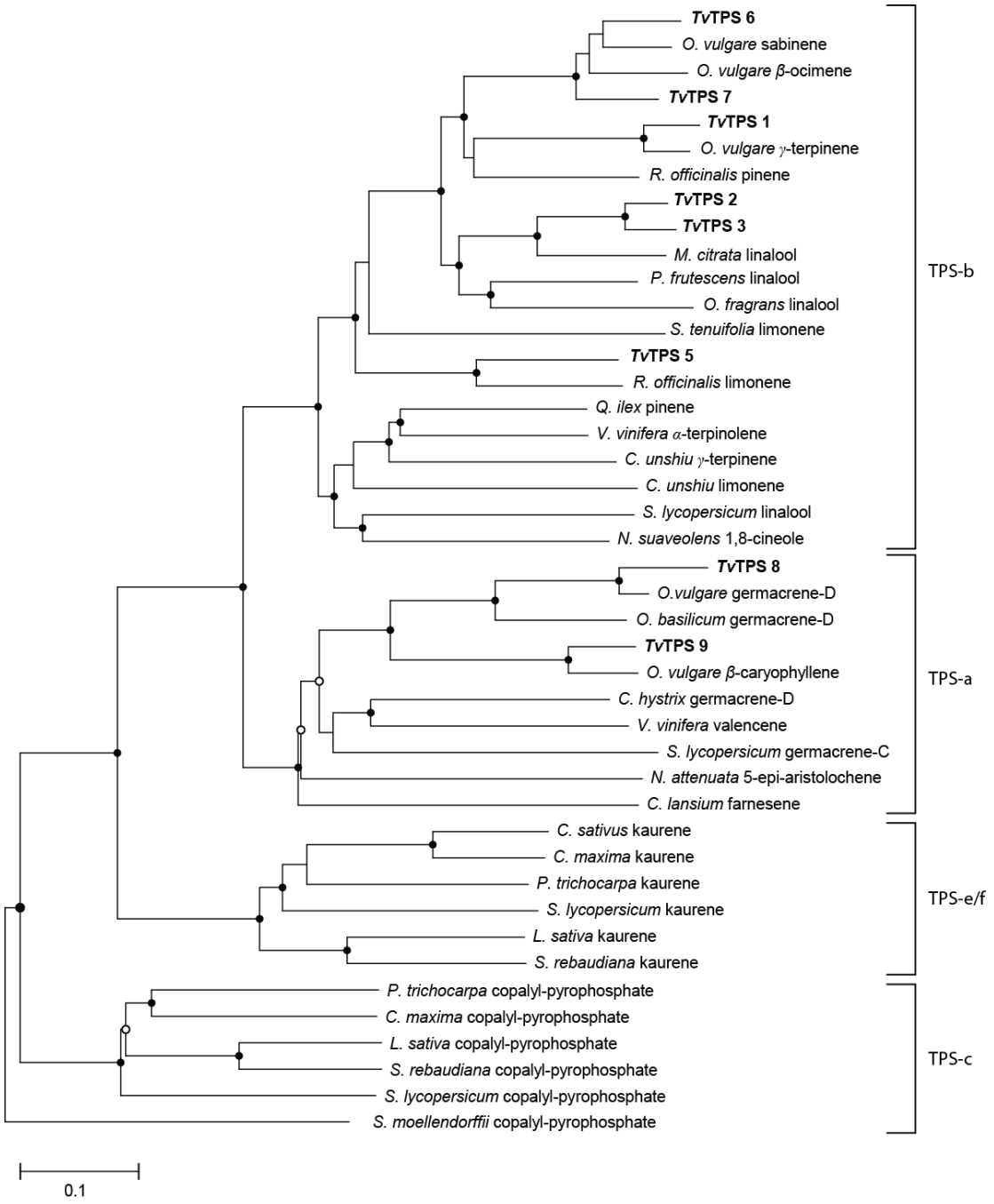


Figure 3-5: Dendrogram analysis showing the positioning of mono- and sesquiterpene synthases isolated from *T. vulgaris* and functionally related terpene synthases in a phylogenetic context. The dendrogram was constructed using the Neighbor-Joining method and rooted with the copalyl-pyrophosphate synthase from the spikemoss *Selaginella moellendorffii* and TPS subfamilies were indicated as defined Chen et al. (2011). Bootstrap values of ≥ 80 % are indicated by black dots on the nodes, while values of ≤ 45 % are indicated by circles on the nodes (10.000 replicates). Sequences determined in this thesis are in boldface. The name of the major enzyme product is given after the

abbreviation of the species. GenBank accession numbers are in parentheses. *C. hystrix – Citrus hystrix*: germacrene D synthase (ADX01384); *C. unshiu – Citrus unshiu*: d-limonene synthase (BAD27257), *γ*-terpinene synthase (BAD27259); *C. lansium – Clausena lansium*: farnesene synthase TPS2-1 (ADR71055); *C. sativus – Cucumis sativus*: ent-kaurene synthase (BAB19275); *C. maxima* - *Cucurbita maxima*: copalyl-pyrophosphate synthase (AAD04292), ent-kaurene synthase (AAB39482); *L. sativa* - *Lactuca sativa*: copalyl-pyrophosphate synthase (BAB12440), ent-kaurene synthase (BAB12441); *M. citrata – Mentha citrata*: linalool synthase (AAL99381); *N. attenuata – Nicotiana attenuata*: 5-epi-aristolochene synthase (AF542544); *N. suaveolens – Nicotiana suaveolens*: 1,8-cineole synthase (ABP88782); *O. basilicum – Ocimum basilicum*: germacrene D synthase (AAV63786); *O. fragrans – Osmanthus fragrans*: linalool synthase (ACM92062); *O. vulgare – Origanum vulgare*: caryophyllene synthase (GU385970), germacrene D synthase (GU385976), *γ*-terpinene synthase (GU385978), sabinene synthase (GU385980), *β*-ocimene synthase (GU385967); *P. frutescens – Perilla frutescens*: linalool synthase (AAL38029); *P. trichocarpa* - *Populus trichocarpa*: copalyl-pyrophosphate synthase (EEE81383), ent-kaurene synthase (EEE88653); *Q. ilex – Quercus ilex*: pinene synthase (CAK55186); *R. officinalis – Rosmarinus officinalis*: limonene synthase (ABD77416), pinene synthase (ABP01684); *S. moellendorffii* - *Selaginella moellendorffii* copalyl-pyrophosphate synthase (EFJ31965); *S. lycopersicum* - *Solanum lycopersicum:* copalyl-pyrophosphate synthase (BAA84918), ent-kaurene synthase (AEP82778), germacrene C synthase (AAC39432), linalool synthase (AEM05855); *S. rebaudiana* - *Stevia rebaudiana*: copalyl-pyrophosphate synthase (AAB87091), kaurene synthase (AAD34295); *S. tenuifolia – Schizonepeta tenuifolia*: limonene synthase (AAG01140); *Tv*TPS – *Thymus vulgaris*: *γ*-terpinene synthase (*Tv*TPS1), linalool synthases (*Tv*TPS2 and *Tv*TPS3), inactive protein (*Tv*TPS5), sabinene hydrate synthases (*Tv*TPS6 and *Tv*TPS7), (*E*)-*β*-caryophyllene synthase (*Tv*TPS9), partial sequence (*Tv*TPS8), *V. vinifera – Vitis vinifera*: valencene synthase (AY561843), *α*-terpinolene synthase (AY572987).

The subfamily TPS-c, which is most conserved among land plants, was defined as an out-group and used to root the tree, which was visualized with TreeView and Adobe Illustrator (Figure 3-5). The phylogenetic tree deliberately only includes sequences from Viridiplantae belonging to the eucotyledons, while other Angiosperm or Gymnosperm sequences were ignored, with the exception of the copalyl-pyrophosphate synthase from the lycophyte *Selaginella moellendorffii*. The tree therefore does not show several subfamilies of the seven TPS families as defined by Chen et al. (2011).

The terpene synthases isolated from *T. vulgaris* fall into two clades thereby separating mono- from sesquiterpene synthases (Figure 3-5). The *T. vulgaris* monoterpene synthases are situated in the TPS-b subfamily, while the sesquiterpene synthases are situated in the TPS-a subfamily.

The monoterpene synthases most closely related to the ones isolated from thyme have all been isolated from members of the Lamiaceae, while sequences of taxonomically more distant plant species are not found in close proximity to the thyme sequences. An enlargement of the TPS-b subfamily makes these subtle differences in the placement of the thyme monoterpene synthases more evident (Figure 3-5 and Figure S1). The Lamiaceae linalool and geraniol synthases, with the exception of the *L. latifolia* linalool synthase and the *O. basilicum* linalool and geraniol synthases, cluster closely together. The clustering of these sequences appears to reflect both catalytic function as well as affiliation to plant families. These dual criteria become especially clear when the positioning of the *γ*-terpinene synthase *Tv*TPS1 is considered. Despite having similar catalytic functions, the *Citrus* sp. *γ*-terpinene synthase sequences are clearly separated from the *γ*-terpinene synthase sequences of *T. vulgaris*, *O. vulgare*, and *O. syriacum*. The two sabinene hydrate synthases *Tv*TPS6 and *Tv*TPS7 too, are clustered together with the *O. vulgare* sabinene and (*E*)-*β*-ocimene synthases, while the sabinene synthases from *S. officinalis* and *S. pomifera* cluster closer together with the *S. officinalis* bornyl diphosphate synthase and the *R. officinalis* pinene synthase.

An enlargement of a dendrogram of the TPS-a subfamily, which comprises the thyme sequiterpene synthases, results in similar observations (Figure 3-5 and Figure S2). However, since the number of sesquiterpene synthases isolated from Lamiaceae or related plant families is limited, the dual criteria for placement in the case of *T. vulgaris* sequences in the phylogenetic tree are not as evident as for the members of the TPS-b subfamily. So form the *T. vulgaris* sequences of *Tv*TPS8 and *Tv*TPS9 subcluster with other Nepetoideae sequences of similar functionality, while other sequiterpene sequences of the Lamiaceae or closely related families are far removed, such as the *O. basilicum* cadinene synthase, the *P. frutescens* valencene synthase, and the *P. cablin* germacrene D synthase. The partial sequence of *Tv*TPS8, thought to be a germacrene D synthase, clusters closely together with the germacrene D synthases of *O. vulgare* and *O. basilicum*, while the *O. basilicum* selinene synthase is slightly further removed. The (*E*)-*β*-caryophyllene synthase *Tv*TPS9 on the other hand, clusters closest to the *O. vulgare* (*E*)-*β*-caryophyllene synthase and the *L. angustifolia* (*E*)-*β*-caryophyllene synthase, while it is slightly removed from the *Tv*TPS8-cluster. It should also be mentioned that the germacrene synthases in the phylogenetic tree are often found in close proximity to (*E*)-*β*-caryophyllene synthases of either the same plant family or close plant relatives. This is true for the sequences of *Vitis vinifera*, *O. vulgare*, *T. vulgaris*, *Solanum lycopersicum*, as well as the Asteroideae sequences. However, it should be pointed

out that the bootstrap values of the tree are in parts of the TPS-a clade low in comparison to the TPS-b clade.

Cytochrome P450s isolated from *T. vulgaris*

P450 sequences isolated from thyme chemotypes L_{48} and T_{28}

It has been surmised that the formation of the phenolic monoterpene alcohols thymol and carvacrol in plants involves the hydroxylation of *γ*-terpinene and *p*-cymene by cytochrome P450 monoterpene hydroxylases (Poulose and Croteau 1978a). It was therefore not surprising to find an EST-fragment with high similarity to the *M. spicata* limonene-6-hydroxylase (CYP71D18), which is involved in carveol biosynthesis in mint, in an EST library of *O. vulgare* cultivar 'f02-04' (Crocoll 2011). As both thyme and oregano are closely related plants, this information could be utilized to isolate putative cytochrome P450 candidates involved in the formation of thymol and carvacrol by colony screening and RACE-PCR.

Two full length sequences could be isolated from the *T. vulgaris* L_{48} and T_{28} chemotypes, which, like the sequences isolated from *O. vulgare*, showed high identity levels to the CYP71D18 of mint. The deduced amino acid sequences of both cDNAs were compared with representatives of the known cytochrome P450 families to assign them to their specific family based on an amino acid identity > 40 % (Nelson et al. 1996). Based on these results, the sequences were annotated as CYP71D179 (*Tv*-P450-L_{48}) and CYP71D180 (*Tv*-P450-T_{28}), respectively. In agreement with the rules of cytochrome P450 nomenclature, the CYP71D179 was included under the same gene name as the CYP71D179 isolated from *O. vulgare*, as their amino acid identity is greater than 97 % (Nelson et al. 1996). Henceforth the sequences isolated from *T. vulgaris* will be referred to as *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} in this thesis.

The protein encoded by the *Tv*-CYP71D179-L_{48} sequence consisted of 496 amino acids and had a calculated molecular mass of 56.3 kDa and a calculated isoelectric point of pI 7.7. The highest homology of the deduced amino acid sequence was to that of the *O. vulgare* CYP71D179 (98 % identity) and the *M. spicata* limonene-6-hydroxylase (72 % identity) (Figure 3-6). The coding region of *Tv*-CYP71D180-T_{28}, in turn, has a length of 496 bp and encodes a protein of 499 amino acids in length. The estimated molecular mass for this protein is 55.8 kDa, with an estimated isoelectric point of pI 7.2. The highest similarity of the deduced amino acid sequence of *Tv*-CYP71D180-T_{28} was to that of the CYP71D178

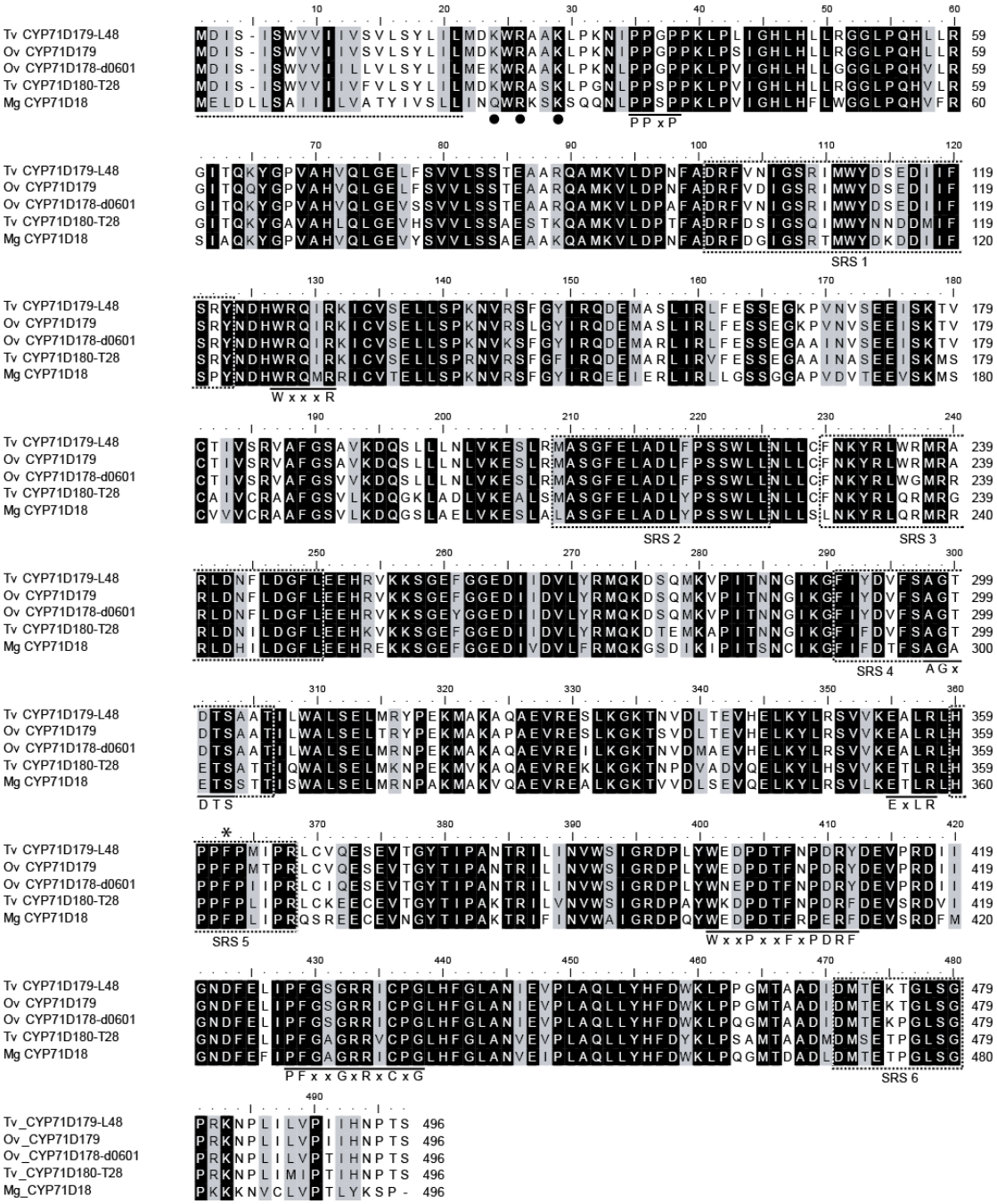
sequence of *O. vulgare* (82 % identity), followed by CYP71D179 from *O. vulgare* (80 % identity), and the limonene-6-hydroxylase from mint (75 % identity). The two putative cytochrome P450 enzymes *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} exhibit an overall identity of 82 % and a similarity of 82 % on the amino acid level, with the regions of greatest variation (Figure 3-6 and 3-7) corresponding roughly to the location of the substrate recognition sites inferred for cytochrome P450 family 2 (Gotoh 1992, Rupasinghe et al. 2003, Crocoll 2011).

Structural analysis of Tv-CYP71D179-L_{48} and Tv-CYP71D180-T_{28} from T. vulgaris.

The deduced amino acid sequences of *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} were compared to the sequences of CYP71D178 and CYP71D179 from *O. vulgare* and the limonene-6-hydroxylase (CYP71D18) from *M. spicata*. Several elements commonly found in microsomal cytochrome P450 proteins could be identified (Figure 3-6). The N-terminal sequences of *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} contain a stretch of about 22 hydrophobic amino acids each, which function as membrane anchor signals and are characteristic for P450 proteins targeted to the endoplasmic reticulum. Just downstream of these hydrophobic amino acids, in positions 24, 26, and 29, are charged residues which serve both as an insertion signal and a **halt-transfer signal** (Sakaguchi et al. 1987). The **PPGP motif** (P-P-[G/S]-P) can be found at residues 35 – 38 in both thyme sequences. This proline and glycine rich region serves as a 'hinge' between the globular domain of the protein and the membrane anchor and is necessary for the proper heme-incorporation and the assembly of functional P450 enzymes in the cell (Yamazaki et al. 1993). The **W-x-x-x-R motif**, located in helix-C and thought to be involved in forming a charged pair with the heme-group, can be found at amino acid residues 127 – 131 in both sequences. The **A-G-x-[D/E]-T-S motif**, required for oxygen-binding, on the other hand is located at residues 298 – 303, while the

Figure 3-6: Alignment of the deduced amino acid sequences of the cytochrome P450 genes. Included are thyme *Tv*CYP71D179-L_{48} (putative P450 enzyme) and *Tv*CYP71D180-T_{28} (putative P450 enzyme) and their nearest homologues, *Ov* CYP71D179, *Ov* CYP71D178-d0601 from *Origanum vulgare*, and limonene-6-hydroxylase *Mg* CYP71D18 (AAQ18706) from *Mentha* x *gracilis*. Amino acids which are common to all five sequences are shaded in black, while residues with similar properties are shaded in grey. Motifs commonly found in P450s are indicated by solid bars, like the proline-rich region (PPxP), the KYG motif, the WxxxR motif, the absolutely conserved ExxR motif, the 'PERF' motif (WxxPxxFxP[D/E]RF), and the heme-binding domain (PFxxGxRxCxG). The hydrophobic membrane anchor domain in the N-terminal region is indicated by a dashed line, while

the halt-transfer signal indicated by acidic amino acids is marked by solid dots. The sequence recognition sites (SRS 1 – 6) are framed by dotted lines. The alignment was created with the ClustalX2 program using the default settings, adjusted by hand in the BioEdit program, and edited in Adobe Illustrator CS.



highly conserved **E-x-L-R motif** is situated at residues 355 – 358. This motif is known to be positioned in the conserved region of P450 sequences and coincides with Helix K in

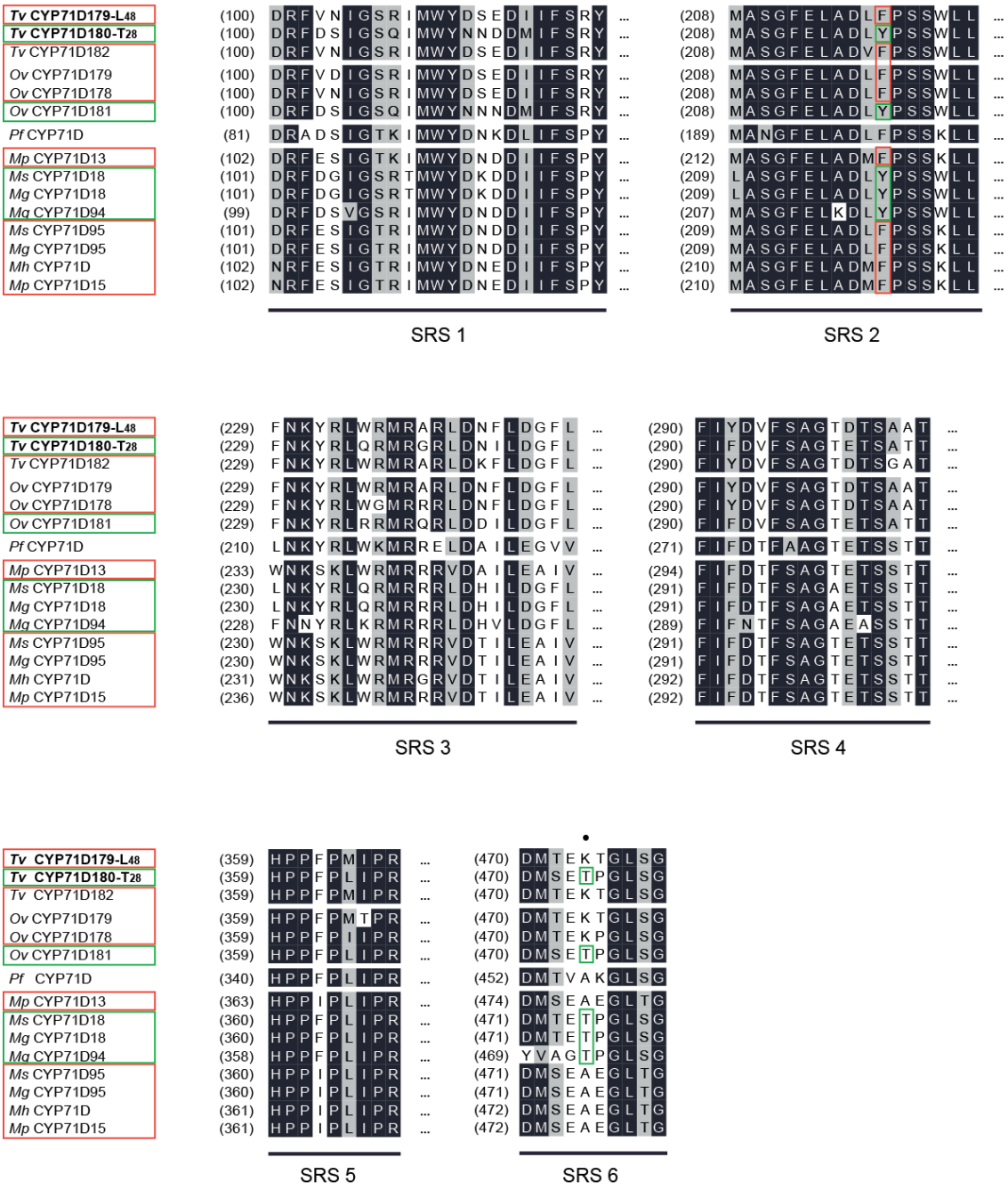
P450cam (Hasemann et al. 1995). A typical aromatic sequence of P450s with three aromatic (A_1 – A_3) residues and two highly conserved proline residues normally designated as A_1-x-x-**P**-x-x-A_2-x-**P**-x-x-A_3, is situated at position 402 – 413 and is present as W-x-x-**P**-x-x-F-x-**P**-[D/E]-R-F. This conserved sequence is thought to interact through a set of salt bridges with the ExLR-motif, thus stabilizing the overall structure (Feyereisen 2012). Finally the P450 enzyme 'signature'-motif, **P-F-x-x-G-x-R-x-C-x-G**, of which the cysteine residue is known to be an important ligand for heme binding and is also responsible for the characteristic 450 nm absorption of the Fe^{II}-CO complex of cytochrome P450s (Feyereisen 2012), is situated at amino acid positions 428 – 438.

Substrate recognition sites of Tv-CYP71D179-L_{48} and Tv-CYP71D180-T_{28}

The so-called substrate recognition sites (SRS) were first inferred by Gotoh (1992), and have shown considerable sequence conservation in the group 2 cytochrome P450s (Rupasinghe et al. 2003, Schuler and Rupasinghe 2011). However, the substrate recognition sites SRS 1, SRS 4, SRS 5, and SRS 6 show relatively high sequence conservation (Rupasinghe et al. 2003), and can be readily recognized without structural modeling (Mansuy and Renaud 1995). However, the location of SRS 2 and SRS 3 can be estimated, even though they contain no overall conserved amino acid residues, since P450s of the group 2 share a high degree of secondary and tertiary structural homology in which secondary structure elements can be found in similar locations (Rupasinghe et al. 2003). The deduction of possible substrate recognition sites in *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} was performed by utilizing alignments of other P450s with known or modeled SRS (Gotoh 1992, Rupasinghe et al. 2003, Crocoll 2011), and using the SWISS-MODEL workspace (Figure 3-7) (Jones 1999, Arnold et al. 2006).

Figure 3-7: Alignment of sequence recognition sites (SRS 1 – 6) for selected CYP71D enzymes of the Lamiaceae. Amino acids common to thirteen or more sequences are shaded in black, while residues with similar properties are shaded in grey. The sequence recognition sites (SRS) are indicated by solid bars. Two positions possibly responsible for the regiospecificity are marked by solid dots. Enzymes with C2- and C6-hydroxylation activity are framed in red, while enzymes with C3-hydroxylation activity are framed in green. *Ov* CYP71D178 is framed in red, although it is able to also perform C2-hydroxylations (Crocoll 2011). GenBank accession numbers are in parentheses. *Mentha* x *gracilis*: *Mg* CYP71D18 (Q6WKZ1), *Mg* CYP71D94 (Q6WKZ0), *Mg* CYP71D95 (Q6WKY9); *Mentha haplocalyx*: *Mg* CYP71D (ABR15423); *Mentha piperita*: *Mp* CYP71D13 (Q9XHE7), *Mp* CYP71D15 (Q9XHE6); *Mentha spicata*: *Ms* CYP71D18 (Q9XHE8), *Ms* CYP71D95 (Q6IV13);

Origanum vulgare: *Ov* CYP71D178, *Ov* CYP71D179, *Ov* CYP71D181; *Perilla frutescens*: *Pf* CYP71D174 (GQ120438); *Thymus vulgaris*: *Tv* CYP71D179-L_{48}, *Tv* CYP71D180-T_{28}.The alignment was created with ClustalX2 using the default settings, while the graphic display was created with BioEdit and Adobe Illustrator CS.



The SRS sequences of *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} are situated as follows: SRS 1 can be found in both sequences at amino acid residues 100 – 122, SRS 2 at

residues 210 – 225, SRS 3 at residues 230 – 250, SRS 4 at residues 290 – 305, SRS 5 at residues 359 – 367, and SRS 6 at residues 470 – 479 (Figure 3-7). The locations of SRS 2 and SRS 3 were determined by their location in the predicted secondary structures (SRS 2 is located at the end of helix-F, while SRS 3 is located at the beginning of helix-G (Gotoh 1992)), since the amino acid conservations in these two substrate recognition sites are too low to be readily identified, as mentioned above. The alignment of the SRS regions of the cytochrome P450 CYP71D sequences studied, shows that SRS 5 (2-3 / 9 differences) and SRS 2 (4-5 / 17 differences) are highly conserved, while SRS 1 (11-12 / 23 differences), SRS 4 (6-7 / 16 differences), and SRS 3 (13-14 / 21 differences) are less so (Figure 3-7). When the SRS regions of *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} are compared to each other the number or different amino acids is even less. SRS 2 and SRS 5 each differ in only one amino acid (Phe-218/Tyr-218 and Met-364/Leu-364, respectively), while SRS 3, SRS 4, and SRS 6 each show three differences, and SRS 1 has again the highest number of differences (7 / 23 differences). However, when the focus of the comparison is not on the individual CYP71D sequences but rather on the regiospecificity of the enzymes, only two amino acids seem to play an important role. These can be found in SRS 2 and SRS 6. *Tv*-CYP71D180-T_{28} and the other C2- and C6-hydroxylation regiospecific sequence possess a tyrosine (Y) in position 218 of *Tv*-CYP71D180-T_{28} and corresponding positions, while the C3- and C7-hydroxylation sequences have a phenylalanine (F) in this position. The C2- and C6-hydroxylation specific sequences possess in the SRS 6 a threonine (T) in position 474, while the C3- and C7-hydroxylation specific sequences possess either a lysine (K) or an alanine (A) in the corresponding position (Figure 3-7).

Evolutionary relationship between CYP71D179-L_{48} and CYP71D180-T_{28} and other members of the CYP71D subfamily

A phylogenetic tree was constructed to illustrate the apparent evolutionary relationship of the CYP71D subfamily and in particular of *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} to the other members of this cytochrome P450 subfamily, by using the deduced amino acid sequences and analyzing them with ClustalX2 and DNAstar. The resulting phylogram, rooted with the studied members of the CYP71A subfamily and the Liliopsida *Zingiber zerumbet* CYP71BB1 as outgroup, revealed two distinct clusters in the lamiid CYP71D subfamily (Figure 3-8).

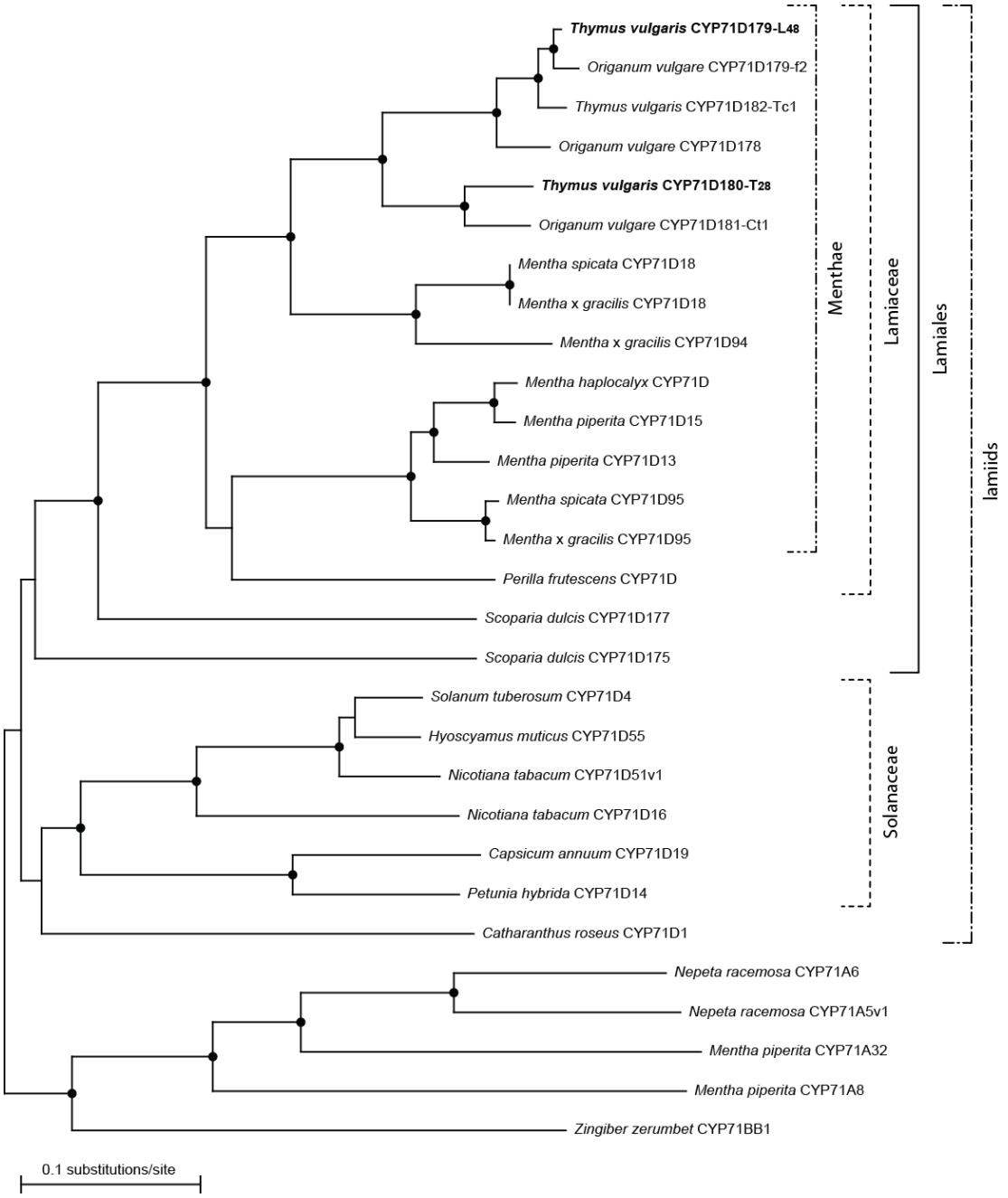


Figure 3-8: Dendrogram analysis of cytochrome P450 sequences from the CYP71D subfamily. The dendrogram was constructed using the Neighbor-Joining method and rooted with the *Zingiber zerumbet* CYP71BB1 sequence. Black dots on the nodes indicate bootstrap values of ≥ 79 % (10.000 replicates). Sequences determined in this thesis are in boldface. GenBank accession numbers are in parentheses. *Capsicum annuum*: CYP71D19 (AAF27282); *Catharanthus roseus*: CYP71D1 (AY192573); *Hyoscyamus muticus*: CYP71D55 (A6YIH8); *Mentha* x *gracilis*: CYP71D18 (Q6WKZ1), CYP71D94 (Q6WKZ0), CYP71D95 (Q6WKY9); *Mentha haplocalyx*: CYP71D

(ABR15423); *Mentha piperita*: CYP71A8 (Q42716), CYP71A32 (Q947B7), CYP71D13 (Q9XHE7), CYP71D15 (Q9XHE6); *Mentha spicata*: CYP71D18 (Q9XHE8), CYP71D95 (Q6IV13); *Nepeta racemosa*: CYP71A5v1 (CAA70575), CYP71A6 (O04164); *Nicotiana tabacum*: CYP71D16 (AAD47832), CYP71D51v1 (DQ350344); *Origanum vulgare*: CYP71D178, CYP71D179-f2, CYP71D181-Ct1; *Perilla frutescens*: CYP71D174 (GQ120438); *Petunia hybrida*: CYP71D14 (AB028462); *Scoparia dulcis*: CYP71D175 (GU592503), CYP71D177 (GU205276); *Solanum tuberosum*: CYP71D4 (AJ296346) *Thymus vulgaris*: CYP71D179-L_{48}, CYP71D180-T_{28}, CYP71D182-Tc1; *Zingiber zerumbet*: CYP71BB1 (AB331235).

The clustering separates Lamiales CYP71D sequences, containing *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28}, from non-Lamiales lamiid CYP71D sequences. The sequence identity inside the non-Lamiales cluster can be as low as 50 % (CYP71D1 to CYP71D16) and as high as 93 % (CYP71D4 to CYP71D55), while the identity level inside the Lamiales cluster shows a very broad range from 43 to 100 %. The Lamiales CYP71D sequences are further divided into Lamiaceae CYP71D and non-Lamiaceae CYP71D, of which the former consists, solely of sequences isolated from members of the Nepetoideae, which share an identity level between 59 and 100 %. The two thyme sequences *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} cluster closest together with CYP71D sequences isolated from *T. vulgaris* and *O. vulgare*, which are thought to use *γ*-terpinene as a substrate (Crocoll 2011). The sequence identity inside this group is comparatively high and ranges from 80 to 98 %. This group forms, together with three CYP71D sequences isolated from *Mentha* sp., a subcluster inside the Nepetoideae cluster. Two of these *Mentha* P450s have been described as (-)-4*S*-limonene-6-hydroxylases, using (-)-4*S*-limonene to produce (-)-*trans*-carveol, by performing a C6-allylic hydroxylation (Lupien et al. 1999, Bertea et al. 2003). The second subcluster inside the Nepetoideae cluster, not containing the thyme and oregano sequences, is formed by five CYP71D sequences also isolated from *Mentha* sp., of which four have been described as limonene-3-hydroxylases producing *trans*-isopiperitenol (Lupien et al. 1999, Schalk and Croteau 2000, Wüst et al. 2001, Lücker et al. 2004a), as well as a putative (-)-limonene-7-hydroxylase described for *P. frutescens* (Mau et al. 2010).

Discussion

In the course of this thesis, several gene sequences from two large families of enzymes involved in terpene biosynthesis, terpene synthases (TPS) and cytochrome P450 monooxygenases (P450), were successfully isolated from *T. vulgaris*. All of the proteins encoded by these sequences are likely to be involved in the production of the essential oil components found in the different chemotypes of thyme. After heterologous expression in *E. coli*, five of the isolated monoterpene synthases were active. Two produced linalool, two produced sabinene hydrates (Krause et al. 2013), and one produced *γ*-terpinene as its main product, the latter of which can be converted further to thymol or carvacrol. Linalool, the sabinene hydrates, as well as thymol and carvacrol are all major accumulated monoterpenes in *T. vulgaris* and define four of the seven known chemotypes in this species. The two major sesquiterpenes, (*E*)-*β*-caryophyllene and germacrene D, found in the essential oil of all thyme chemotypes, are produced by two separate sesquiterpene synthases of which one could be successfully isolated and expressed.

Of the active monoterpene synthases isolated, three can be described as forming multiple products. In addition to a single main product, these synthases form significant amounts of additional products. Nearly half of all mono- and sesquiterpene synthases characterized can be defined as multi-product enzymes in which side products represent at least 10 % of the total product spectrum (Degenhardt et al. 2009).

In addition to these terpene synthase genes, two cytochrome P450 genes could be isolated which are thought to catalyze the transformation of *γ*-terpinene to the phenolic monoterpene alcohols thymol and carvacrol, the two major accumulated phenolic monoterpenes in *T. vulgaris*.

Terpene synthases isolated from *T. vulgaris* are closely related to other Lamiaceae terpene synthases

Comparison of the isolated TPS from thyme to other known angiosperm TPS, based on sequence similarity on the amino acid level made it evident that they are closely related to previously described TPS from other Lamiaceae species. This was expected, as previous studies often demonstrated that TPS from a certain taxon are generally more similar to TPS from related taxa than to other TPS from other angiosperms producing the same product (Iijima et al. 2004a, Landmann et al. 2007, Crocoll et al. 2010). The monoterpene synthases

Tvtps1, *Tvtps2*, *Tvtps3*, *Tvtps5*, *Tvtps6*, and *Tvtps7* clustered in the TPS-b subfamily (Chen et al. 2011), in close proximity to other Lamiaceae monoterpene synthase sequences, while the sesquiterpene synthase *Tvtps9* and the partial sequence of the putative sesquiterpene synthase *Tvtps8* could both be found in close proximity to members of the TPS-a subfamily (Figure 3-5).

Sequence similarity places the *T. vulgaris* sequences closest together with TPS sequences of phyletically close members of the Lamiaceae, the members of the three Nepetoideae tribes Mentheae (*Mentha*, *Salvia*, *Thymus*, *Origanum*), Elsholtzieae (*Perilla*), and Ocimeae (*Ocimum*, *Lavandula*) in particular (Figure 3-5, Table 3-1, and Table 3-2). For example the *T. vulgaris* γ-terpinene synthase *Tv*TPS1 can be found in close proximity to the Menthae γ-terpinene synthases from *O. vulgare* (Crocoll et al. 2010) and *O. syriacum* (direct submission to NCBI-database). The only other known γ-terpinene synthases were isolated from the Rutaceae members *Citrus unshiu* (Shimada et al. 2004) and *C. limon* (Lücker et al. 2002), which are phyletically far removed from the Lamiaceae clustering with other members of the rosids, and in particular members of the Sapindales. Similarly, the two sabinene-hydrate synthases *Tv*TPS6 and *Tv*TPS7 are found in close proximity to the *O. vulgare* sabinene synthase and *β*-ocimene synthase (Crocoll et al. 2010) from the Menthae, while the putative monoterpene synthase *Tv*TPS5 (activity not demonstrated) clusters together with another Lamiaceae TPS, the *Rosmarinus officinalis* limonene synthase (direct submission to NCBI-database).

The placement of the *T. vulgaris* monoterpene synthase sequences is thus generally congruent with the currently accepted taxonomic relationships (Petersen et al. 2009). However the placement of several Nepetoideae linalool synthases, are difficult to explain. Most Nepetoideae linalool synthases cluster together with the two linalool synthases *Tv*TPS2 and *Tv*TPS3, but the *Ocimum basilicum* linalool synthase (Iijima et al. 2004a) and the putative *Lavandula latifolia* linalool synthase (direct submission to NCBI-database) are exceptions (Figure 3-5 and Figure S1). The *O. basilicum* linalool synthase, as well as the *O. basilicum* geraniol synthase (Iijima et al. 2004b) cluster together with linalool and geraniol synthases of the phyletically distant *Vitis vinifera* (Vitaceae) and *Phyla dulcis* (Verbenaceae), respectively. The *L. angustifolia* linalool synthase (Landmann et al. 2007) clusters with the linalool and geraniol synthases of phyletically more distant *Mentha*, *Perilla*, and *Thymus* species rather than with the other Ocimeae linalool and geraniol synthases. This suggests that linalool and geraniol synthases may have arisen independently on multiple occasions.

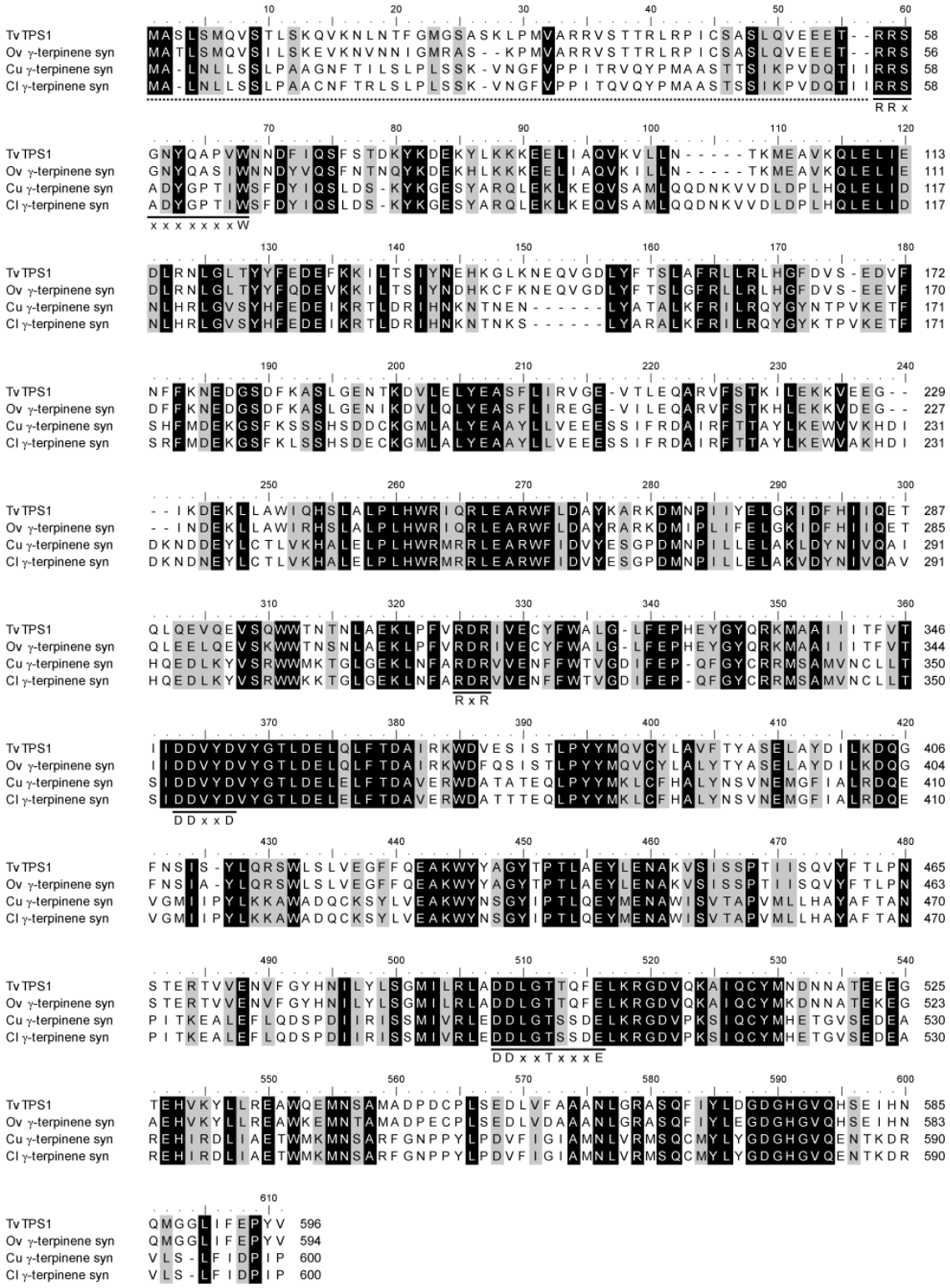


Figure 3-9: Deduced amino acid comparison of *Tv*TPS1 to *γ*-terpinene synthases from *Origanum* and *Citrus*. Alignment of *Thymus vulgaris Tv*TPS1 (*γ*-terpinene synthase) predicted protein with the functionally similar *γ*-terpinene synthases from *Origanum vulgare* (GU385978), *Citrus unshiu*

(BAD27259), and *Citrus limon* (AAM53943), respectively. Residues shaded in black indicate conserved residues in all four sequences, while residues with similar properties are shaded in grey. The RRx_8W motif, the RxR motif, as well as the metal binding motifs DDxxD and NSE/DTE, which are highly conserved in plant terpene synthases are indicated by solid bars. The alignment was created with the ClustalX2 program using the default settings, adjusted by hand in the BioEdit program, and edited in Adobe Illustrator CS.

The amino acid identities among monoterpene synthases from Lamiaceae species are generally rather high, irrespective of their catalytic functions, and usually range between 55 – 90 %, while the identity level to monoterpene synthases isolated from species outside this plant family are comparatively low (Table 3-1). For instance the overall identity between the *T. vulgaris* γ-terpinene synthase *Tv*TPS1 and the *O. vulgare* γ-terpinene synthase is 91.5 %, while the overall identity of *Tv*TPS1 shared with the γ-terpinene synthases from *C. limon* and *C. unshiu* is only 42.4 and 42.6 %, respectively. The *Citrus* sequences in turn share an overall amino acid identity of 98 % with each other. The identity levels among these four genes are considerably higher however, when only amino acid regions involved in the formation of the active center are considered (Figure 3-9). This clear separation of the γ-terpinene synthases from Rutaceae and Lamiaceae is a strong indication that these genes are the result of repeated evolution, rather than the result of a shared γ-terpinene synthase ancestor. This is congruent with the general assumption that monoterpene synthases share a general monoterpene synthase ancestor but may not necessarily share ancestors generating chemically similar products (Gang 2005).

Since the number of sesquiterpene synthases described for members of the Lamiaceae is rather limited, the positioning of the sequences isolated from *T. vulgaris* is not as precise as for the Lamiaceae monoterpene synthases. Still, it is evident that the sesquiterpene synthases too cluster according to the general taxonomic relationships of the plants that produce them (Figure 3-5).

T. vulgaris terpene synthases also cluster with terpene synthases having similar catalytic functions

When viewed at lower taxonomic levels, the terpene synthases isolated from *T. vulgaris* exhibit tendencies to cluster with other TPS with similar product specificity. For instance, the two linalool synthases *Tv*TPS2 and *Tv*TPS3, have a homology between 79.2 and 66.9 % on the amino acid level to the linalool and geraniol synthases of *Mentha* sp., *Perilla*

sp., and *L. latifolia*, while their identities to other *T. vulgaris* synthases range between only 64.5 and 49.7 % (Table 3-1). The same observation is true for the thyme *γ*-terpinene synthase *Tv*TPS1, both sabinene-hydrate synthases *Tv*TPS6 and *Tv*TPS7, as well as the two sesquiterpene synthases *Tv*TPS8 and *Tv*TPS9, which all show higher identity levels to functionally identical synthases from other plant species than to intraspecific functionally different synthases from the same species (Figure 3-5, Table 3-2, and Figure S2).

Similar to the monoterpene synthases, the *T. vulgaris* sesquiterpene synthases also show a high similarity to enzymes with like catalytic functions. The sesquiterpene synthase *Tv*TPS9, an (*E*)-*β*-caryophyllene synthase, shows a very high similarity (88.4 %) to the *O. vulgare* (*E*)-*β*-caryophyllene synthase (Crocoll et al. 2010) and a relatively high similarity (68.7 %) to the *L. angustifolia* (Jullien et al. 2014) (*E*)-*β*-caryophyllene synthase, while it shares similarities of only 51.2 % and 58.3 % with the other described Lamiaceae sesquiterpene synthases. This apparent correlation of sequence similarity and enzymatic activity in Lamiaceae terpene synthases enables the guarded prediction that *Tv*TPS8, isolated only as a partial sequence in this study, is in fact a germacrene D synthase. The partial sequence of *Tv*TPS8 is 88.0 and 74.7 % similar, to the known germacrene D synthases from *O. vulgare* (Crocoll et al. 2010) and *O. basilicum* (Iijima et al. 2004a) of the Lamiaceae family, respectively. The similarity *Tv*TPS8 shares with *Tv*TPS9 and other Lamiaceae sesquiterpene synthases however is quite low ranging between 50.2 and 56.5 % (Table 3-2).

Sequence motifs and common elements found in terpene synthases

Most plant mono- and sesquiterpene synthases belong structurally to the large group known as class I terpenoid synthases based on general structural organization (Aaron and Christianson 2010). Class I TPS have two structural domains, the N-terminal domain, whose function in catalysis is currently unknown, and the C-terminal domain containing the catalytic center (Starks et al. 1997, Aaron and Christianson 2010). The monoterpene synthases, including the inactive *Tv*TPS5, as well as the sesquiterpene synthases isolated from thyme in this study all belong to the class I type of TPS, as do all other Lamiaceae TPS characterized so far.

When the primary structures of the mono- and sesquiterpene synthases from *T. vulgaris* are aligned with those of other members of the Lamiaceae and other angiosperm plant families, similar characteristics become evident. The RRx_8W, the $RxRx_6W$, the DDxxD, and the NSE/DTE motifs can be found in most of them and in all of the sequences isolated

from thyme (Figure 3-1 and Figure 3-3). The two aspartate rich motifs, DDxxD and NSE/DTE, are essential for terpene biosynthesis and their role in the reaction mechanism has been well studied. While the aspartates of the DDxxD motif and the NSE/DTE motif are essential to coordinate the trinuclear metal cluster involved in the ionization steps in terpenoid biosynthesis, the RRx_8W motif and the $RxRx_6W$ motif are thought to be involved in terpene cyclization (Aaron and Christianson 2010). In addition to these motifs, which are directly involved in the catalytic activity of the enzymes, several other elements exist in one or both terpene synthase families that seem to be only indirectly involved in catalytic activity. For example, the '[A/S/T]LxF[R/K][L/I][L/V/F]R' and the '[L/V]xLYEA[S/T/A]x6[E/D]' elements are found in mono- and sesquiterpene synthases of the angiosperms, while the 'GVYxEP' element seems to be specific to angiosperm sesquiterpene synthases (Cai et al. 2002). The two former elements are situated in the N-terminal domain of the enzyme and are therefore likely to be essential for the overall structure of the enzyme rather than being directly involved in catalysis. The latter element is situated in the C-terminal domain of the enzyme and its function remains unknown.

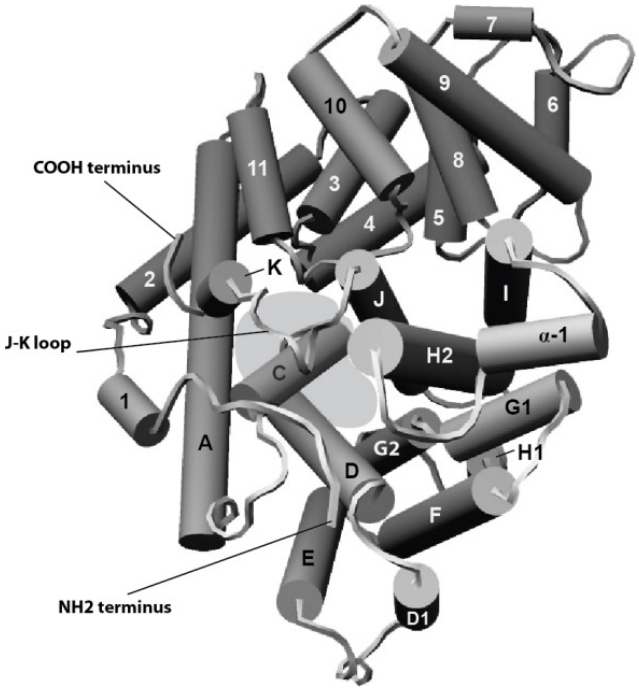


Figure 3-10: Schematic representation of *T. vulgaris Tv*TPS1 three-dimensional structure. The COOH-terminus and NH_2-terminus are indicated, as is the J-K loop. The structure elements were named in accordance to Starks et al. (1997), with the C-terminal domain in the foreground and the N-terminal domain in the background. The active center is indicated as a light grey cloud. The illustration

was created using the SWISS-MODEL workspace, with the 3D-structure of the *Salvia officinalis* bornyl diphosphate synthase (Whittington et al. 2002) as template, visualized with PyMOL, and subsequently edited in Adobe Illustrator CS.

The conserved structural similarities are not confined to the primary structure of the enzymes, but can also be found in the secondary structure. The active center in type I terpenoid synthases, for example, is comprised of six α-helices (C, D, F, G, H, and J), while the ceiling is formed by the A-C loop and the J-K loop (Starks et al. 1997) (Figure 3-10). The trinuclear metal cluster, which is essential for the ionization steps in terpenoid biosynthesis, is coordinated by the DDxxD motif, situated on helix-D, and the NSE/DTE motif, situated on helix-H, which are facing each other and are located at the entrance of the active pocket (Aaron and Christianson 2010). The entrance of the catalytic pocket itself is mainly comprised of polar amino acids, which are thought to bind the ionic part of the substrate, while the aliphatic remainder of the substrate reaches deep into the pocket (Starks et al. 1997). The lower parts of the enzymatic pocket in turn, are dominated by hydrophobic side chains.

The as-yet-undiscovered thyme geraniol synthase is probably similar to the linalool synthase already described

The acyclic monoterpenes geraniol and linalool define two of the seven *T. vulgaris* chemotypes. Both monoterpene alcohols can be formed by nucleophilic attack of water on the respective cation, before the dissociated diphosphate moiety can reattack or the cyclization of the α-terpinyl cation can occur (Figure 3-11). Geraniol and linalool synthases are in this regard considered as 'defective' forms of terpene cyclases. This is a logical explanation for the apparent independent development of both enzymes from very diverse parental terpene cyclases in the different TPS subclasses (Crowell et al. 2002, Landmann et al. 2007).

Linalool synthases of the Lamiaceae plant family possess a characteristic structural feature. Over ten years ago, a comparison of several terpene synthases by modeling demonstrated that a *M. citrata* linalool synthase differed from the others in possessing a three amino acid deletion in the active site pocket (Crowell et al. 2002), thereby causing a dislocation of the so-called J-K loop and providing an easier access of water during substrate ionization (van Schie et al. 2007). This facilitated water access is believed to result in the premature quenching of the intermediate monoterpenol carbocation, before cyclization to a monoterpene olefin can occur (Figure 3-11). Indeed, amino acid deletions near the C-

terminus seem to play an important role in the determination of the enzymatic properties of several monoterpene synthases of the Lamiaceae plant family (Figure 3-2).

Including recent work, linalool synthases have now been isolated and characterized from four (TPS-b, TSP-e/f, TPS-d, and TPS-g) of the seven subfamilies of terpene synthases, the majority of which are highly specific, producing only linalool (Crowell et al. 2002, Landmann et al. 2007, Chen et al. 2010, Chen et al. 2011). Almost all described Lamiaceae linalool synthases, including *Tv*TPS2 and *Tv*TPS3, show a similar three amino acid deletion in the C-terminal region, as was described above. The sole exceptions to this, as stated earlier, are the linalool synthases from *L. latifolia* and *O. basilicum*. The former sequence is a direct submission to the NCBI database whose functional characterization has not been reported, and may not actually represent a true linalool synthase but possibly a limonene synthase, as inferred from sequence comparison to *L. angustifolia* limonene synthase. The other exceptional linalool synthase from *O. basilicum* clusters with other members of the TPS-g subfamily, a group that does not contain other representatives from the Lamiaceae except the *O. basilicum* geraniol synthase. In this regard it should be mentioned that the terpene synthase sequences of *O. basilicum* frequently differ considerably from other Lamiaceae sequences.

Geraniol synthases carry out reactions similar to linalool synthases involving nuclecophilic attack on acyclic carbocations, but no amino acid deletion can be observed in the Lamiaceae TPS-b clustering geraniol synthase sequences isolated so far (Masumoto et al. 2010). In contrast to the *Perilla* sp. geraniol synthases clusters the *O. basilicum* geraniol synthase (Iijima et al. 2004b) in close proximity to the *O. basilicum* linalool synthase and thus in the TPS-g subcluster (Figure S1). However, it is interesting that the *Ocimum* linalool and geraniol synthases like the *Perilla* sp. linalool and geraniol synthases, despite belonging to different TPS subclusters, share high sequence similarities with each other. This observation supports the assumption that linalool and geraniol reaction mechanisms are very similar which in turn are reflected in amino acid profile conservation. The identity levels *Perilla* sp. geraniol and linalool synthases and *O. basilicum* geraniol and linalool synthases share, are very high, and range between 71 - 81 % and 82 %, respectively. The identity levels the *O. basilicum* geraniol synthase shares with the geraniol synthases from the different *Perilla* species are about 30 %, while the identity shared with geraniol synthases of phyletically further removed sequences are among others 67 % (Lamiales: *Lippia dulcis*, also known as *Phyla dulcis* (Trevir.) Moldenke (Yang et al. 2011)) and 66 % (Gentianales: *Catharanthus*

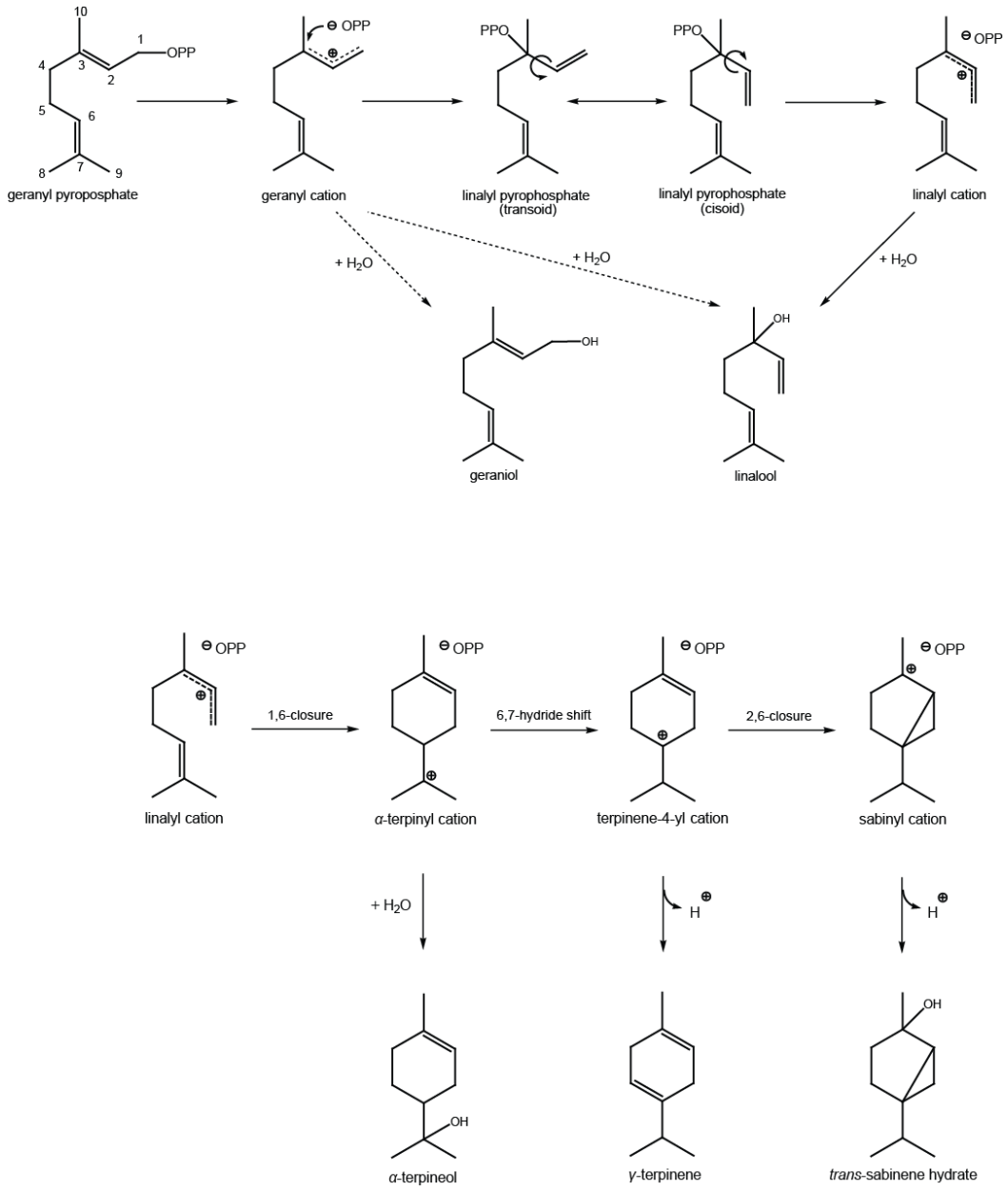


Figure 3-11: Reaction mechanism of pivotal cyclic and acyclic monoterpenes found in *T. vulgaris* chemotypes. The reactions start with the ionization of the geranyl pyrophosphate. Nucleophilic attack of water on the geranyl cation, before the reattack of the dissociated diphosphate moiety (OPP) can occur, leads to the formation of the acyclic monoterpene geraniol, while the formation of the acyclic monoterpene linalool can proceed either via the geranyl cation or via the linalyl cation. The initial cyclization of the linalyl cation to the α-terpinyl cation is required for the formation of the cyclic monoterpenes. The α-terpinyl cation can undergo a range of cyclisations, hydride shifts, and rearrangements before the reaction is terminated either by deprotonation or water capture.

roseus (Simkin et al. 2013)). It can be therefore inferred that the *T. vulgaris* geraniol synthase, not yet identified to date, will most likely share a high identity with the two linalool synthases *Tv*TPS2 and *Tv*TPS3, which are about 70 % identical to the known *Perilla* linalool synthases and about 65 % identical to the known *Perilla* geraniol synthase (Table 3-1), as well as lack the amino acid deletion in the J-K loop of the enzyme similar to the *Perilla* geraniol synthases.

Comparison of the γ-terpinene synthases of thyme, oregano, and citrus

The monoterpenoid γ-terpinene is widely distributed in the plant kingdom and can be found in several plant species belonging to different plant families. Plant genera noted for high γ-terpinene content include *Citrus* (Rutaceae), *Melaleuca* (Myrtaceae), *Origanum* (Lamiaceae: Menthae), and *Thymus* (Menthae). As stated above, γ-terpinene synthases have successfully been isolated from *C. limon* (Lücker et al. 2002), *C. unshiu* (Shimada et al. 2004), *O. vulgare* (Crocoll et al. 2010), and *T. vulgaris*, as described in this thesis. Based on the low similarity between the representatives of different families (less than 45 %) and their origin from phylogenetically distant taxa, the γ-terpinene synthases seem to be the product of convergent evolution. Yet they have some intriguing similarities. While all of them are multiproduct monoterpene enzymes, they synthesize similar amounts of the main product γ-terpinene (71 – 89 %), along with several minor products, such as α-terpinene, limonene, and α-thujene, whose biochemical properties closely resemble each other (Poulose and Croteau 1978b, Lücker et al. 2002, Shimada et al. 2004, Crocoll et al. 2010). Thus despite the lack of much sequence similarity between the γ-terpinene synthases from Rutaceae and Lamiaceae, their active center regions and their product spectrum differ surprisingly little (Figure 3-9). It is plausible to assume that there are basic biochemical and stereochemical requirements for the formation of γ-terpinene by TPS catalysis and these inevitably result in the formation of certain by-products. In future this proposition could be tested further by site-directed mutational studies.

Cytochrome P450 sequences isolated from *T. vulgaris* show a high similarity to Mentheae cytochrome P450 sequences involved in essential oil biosynthesis

Terpenes can undergo additional modifications which are catalyzed by a series of different enzymes, such as dehydrogenases, reductases, methyl transferases, or cytochrome P450 hydroxylases, thereby adding to the diversity of terpenes. Various studies have described the roles P450s can play in the essential oil formation and its regulation in the plant

kingdom. The best studied examples are a number of limonene hydroxylases which catalyze hydroxylations at different carbon atoms of this monoterpene and whose involvement in monoterpene biosynthesis has been emphatically demonstrated (Lupien et al. 1999, Schalk and Croteau 2000, Mau et al. 2010). Different limonene hydroxylases are responsible in peppermint (*Mentha* x *piperita*) and spearmint (*M. spicata*) for the formation of the intermediates of the plant specific essential oil components mentol and carveol. The limonene

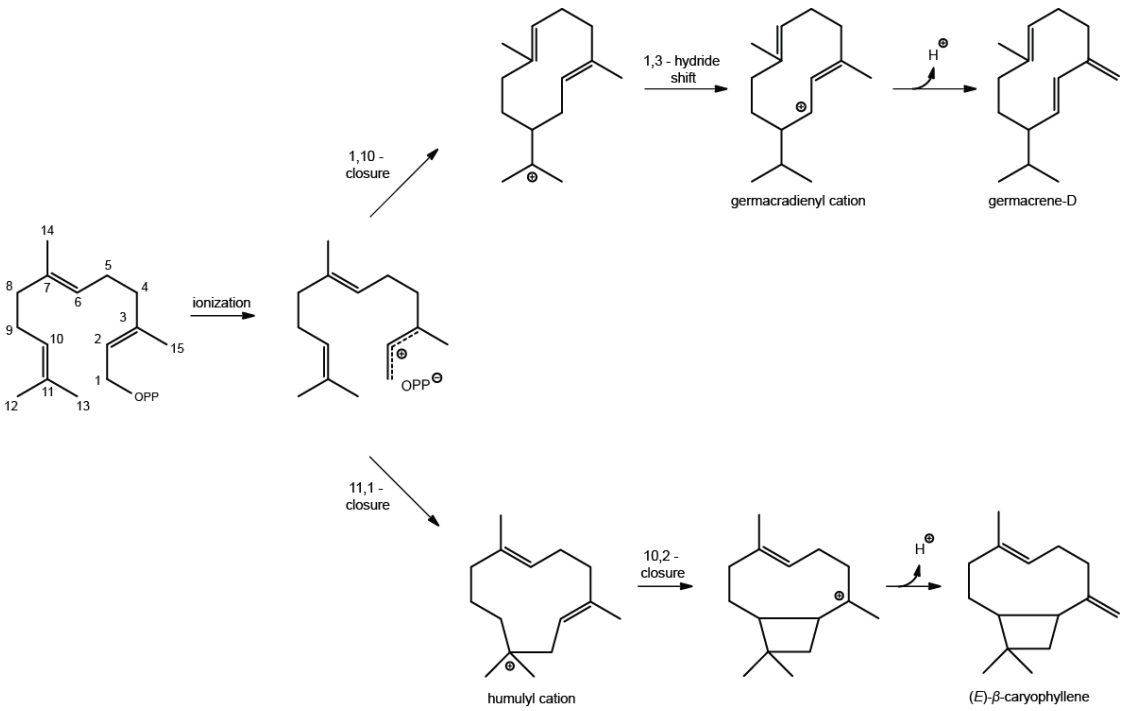


Figure 3-12: Reaction mechanism of pivotal sesquiterpenes found in *T. vulgaris* chemotypes. The reactions start with the ionization of the farnesyl diphosphate. A series of cyclisations and hydride shifts lead to germacrene D and (*E*)-*β*-caryophyllene via the germacradienyl cation and the humulyl cation, respectively. OPP denotes the diphosphate moiety.

hydroxylases CYP71D13 and CYP71D15 in peppermint catalyze a C3-hydroxylation to form (-)-*trans*-isopiperitenol from (-)-*S*-limonene, while the spearmint limonene hydroxylase CYP71D18 catalyzes a C6-hydroxylation to form (-)-*trans*-carveol of said substrate (Karp et al. 1990, Lupien et al. 1999). A further limonene hydroxylase CYP71D174, also from the Lamiaceae plant family, has been described for *P. frutescens*; this enzyme is able to catalyze the hydroxylation of (-)-*S*-limonene at three different position, to either form (-)-*trans*-isopiperitenol (C3-hydroxylation), (-)-*trans*-carveol (C6-hydroxylation), or (-)-perillyl alcohol (C7-hydroxylation) (Mau et al. 2010). In caraway (*Carum carvi*), although a member

of the Apioideae and therefore not closely related to the Lamiaceae, it is assumed that the biosynthesis of (+)-*trans*-carveol is analogous to the biosynthesis of (-)-*trans*-carvone in spearmint (Bouwmeester et al. 1998).

In the late 1970's Poulose and Croteau (1978a) postulated that enzymatic hydroxylation is involved in the thymol and carvacrol biosynthesis in *T. vulgaris*. The hypothesized biosynthetic pathway of thymol and carvacrol in *T. vulgaris* started with *γ*-terpinene (a product of *Tv*TPS1) as the initial monoterpene substrate and proceeded via the aromatic *p*-cymene as an intermediate to thymol and carvacrol, respectively. This hypothesis was supported by the fact that most plant species, which comprise several plant families, which produce thymol or carvacrol, also contain *γ*-terpinene and *p*-cymene. However, it is now believed that *p*-cymene is not an intermediate but a byproduct (Crocoll 2011). It was therefore most likely that *γ*-terpinene was converted to thymol or carvacrol with the help of one or more cytochrome P450 hydroxylases, carrying out hydroxylations comparable to the ones described for the limonene-hydroxylases (Figure 3-13)(Crocoll 2011). As both cyotchrome P450 hydroxylase candidates *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} were isolated from thyme chemotypes with relatively high thymol and carvacrol amounts found in their essential oil, it is very likely that they are involved in the biosynthesis of one or both monoterpene alcohols.

The deduced thyme hydroxylase candidates *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} belong to the CYP71 clan. This clan comprises the largest set of P450s in plants and encompasses next to the CYP71 family among others the CYP79 family of which some of its members are also involved in the synthesis of cyanogenic glucosides and glucosinolates, two other defense-related secondary metabolites (Nelson et al. 2004). In general the cytochrome P450 monooxygenase enzyme family is known for its high sequence diversity, and its limited number of conserved residues found on the sequence level. It is therefore noteworthy that the Lamiaceae CYP71D subgroup displays an unexpected overall high sequence identity. *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} are 81.7 % identical (and 92.5 % similar) to each other. While *Tv*-CYP71D179-L_{48} is 98.2 and 97.4 % identical (and 99.0 % similar) to the *O. vulgare* CYP71D179 and the *T. vulgare* CYP71D182, respectively; is the sequence *Tv*-CYP71D180-T_{28} 94.0 and 92.5 % identical (and 82.9 and 96.8 % similar) to the *O. vulgare* CYP71D members 178 and 181, respectively. Even though both, *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28}, score their highest identity and similarity

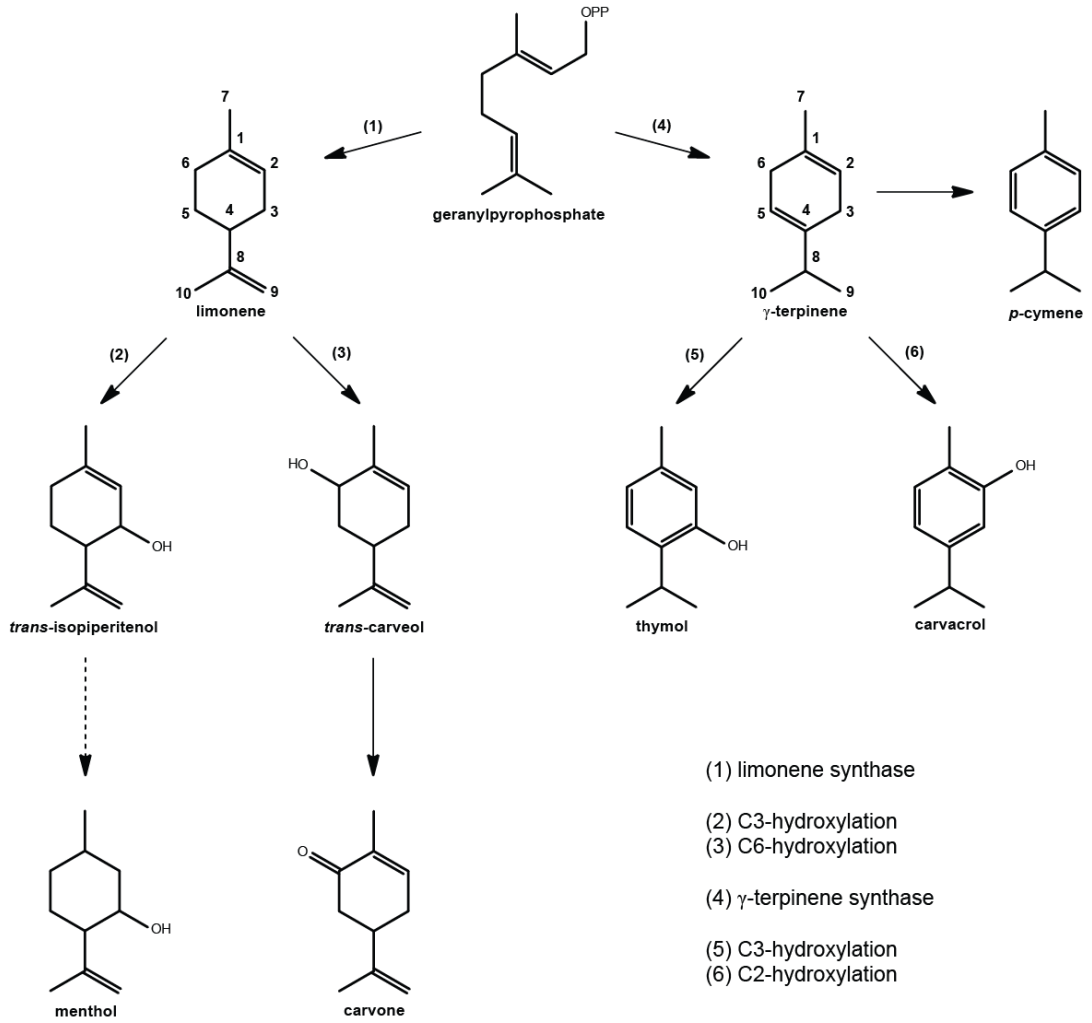


Figure 3-13: Hydroxylations of limonene and *γ*-terpinene in selected members of the Lamiaceae. In peppermint (*Mentha* x *piperita*), limonene is converted to *trans*-isopiperitenol by C3-hydroxylation; in spearmint (*Mentha spicata*), limonene is converted to *trans*-carveol by C6-hydroxylation. In thyme (*Thymus vulgaris*) and oregano (*Origanum vulgare*), *γ*-terpinene is converted to thymol by C3-hydroxylation and carvacrol by C2-hydroxylation. Modified after Crocoll (2011)

percentages with CYP71D members isolated from either thyme and/or oregano, the identity levels to other Lamiaceae members are mostly above 60 % and can go as high as 73 %, while the similarity levels are even higher and range between 80 and 89 %. Sequences isolated from either thyme or oregano cluster with mint limonene-6-hydroxylases, while the mint limonene-3-hydroxylases form another group (Figure 3-8). Nonetheless, the high degree of sequence homology among the Lamiaceae CYP71D members, from six different species of four

different genera, is support for the hypothesis that the members of this subclass arose from a common ancestral cytochrome P450 gene (Nelson et al. 2004, Mau et al. 2010).

Although the primary sequence of P450s of the group 2 type may share very low identity levels; the secondary and tertiary structure levels of most P450 monooxygenases have maintained a surprising amount of conservation. The core structure of the conserved P450-fold is formed by 11 α-helices (named A - K) and four β-sheets (named 1 – 4) (Schuler and Rupasinghe 2011). The substrate recognition sites (SRS), comparatively short regions which are comprised in the catalytic site of most P450s, are involved in the control of substrate access, binding, and catalysis (Gotoh 1992, Schuler and Rupasinghe 2011). SRS 1 corresponds to the highly variable B-C loop region, positioned between helix-B and helix-C, over the catalytically essential heme, SRS 2 is located at the C-terminal end of helix-F, while SRS 3 is located at the N-terminal region of helix-G, SRS 4 corresponds to the N-terminal region of the I-helix extending over the heme, and SRS 5 and SRS 6 correspond to the amino-terminus of β-sheet 1-4 and β-turn at the end of β-sheet 4-1 that both protrude into the catalytic site (Gotoh 1992, Sirim et al. 2010, Schuler and Rupasinghe 2011). The access of the substrate to the binding pocket is limited by the flexible regions in the entrance channel created by helix-F and helix-G, which undergo strong conformational changes upon substrate binding (Poulos and Meharenna 2007, Sirim et al. 2010). The regions directly delimiting the substrate binding pocket and thereby limiting the access of the substrate to the catalytically important heme are the B-C loop region, helix-I, and SRS 5 (Schalk and Croteau 2000, Sirim et al. 2010).

Alignments and structural modeling of the Lamiaceae CYP71D sequences, including *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28}, revealed the possible positioning of the substrate recognition sites 1 – 6, thus making the prediction of likely amino acids responsible for regio- and/or substratespecificity possible (Figure 3-6 and Figure 3-7). The SRS of the Lamiaceae CYP71D appear to be overall well conserved, with SRS 2 (71 % identical) and SRS 5 (66 % identical) being the most conserved, and SRS 3 (33 % identical) and SRS 6 (30 % identical) being the least conserved. When the known catalytic activities are taken into account and the regions are assessed accordingly, the number of identical amino acid residues rises dramatically suggesting a strict conservation of residues essential for substrate selectivity and regiospecificity. The mint C3-hydroxylases for example are between 91 % (SRS 1 and SRS 3) and 100 % (SRS 4, SRS 5, and SRS 6) identical in their sequence

recognition sites. For all sequence recognition sites in P450s, examples can be found in the literature, where site-directed mutagenesis affected the substrate selectivity, the regiospecificity, or the activity.

SRS 5, which delimits part of the substrate pocket and starts shortly after the highly conserved ExxR motif, has been repeatedly shown to play a key role in the selectivity, specificity, and activity of P450s. Seifert and Pleiss (2009) analyzed the SRS 5 region of more than 6300 P450 sequences and could draw several interesting conclusions. They discovered a heme-interacting residue (histidine or arginine), which was most often found 10 or 11 residues after the ExxR motif. In all Lamiaceae CYP71D sequences an arginine can be found in position 10 after the ExxR motif, and thus at the very end of the predicted SRS 5 (Figure 3-7). In addition to this heme-interacting residue they discovered two more protruding residues in the SRS 5 region pointing in the heme-direction in a large part of the assessed sequences. The former was observed more often and could be found at position 5 after the ExxR motif, while the latter was observed at either position 7, 8, or 9 after said motif. In positions 4 and 6 after the ExxR motif, as is also the case in all Lamiaceae CYP71D, they observed a proline in about 40 and 50 % of the analyzed sequences, respectively. Position 5 is predominantly occupied by hydrophobic residues and either isoleucine or phenylalanine can be found in the Lamiaceae CYP71D sequences in this position. Several structure-function studies have confirmed the importance of the residue at position 5 after the ExxR motif for specificity and selectivity (Schalk and Croteau 2000, Takahashi et al. 2007, Seifert and Pleiss 2009, Schuler and Rupasinghe 2011). In particular interesting for this thesis are the mutagenesis studies on spearmint and peppermint limonene hydroxylases performed by Schalk et al. (2000). They could demonstrate that a F363I mutation in the spearmint C6-hydroxylase (CYP71D18) led to a complete conversion to a C3-hydroxylase, while the reciprocal I364F modification in the peppermint C3-hydroxylase (CYP71D15) lead to a properly folded but inactive enzyme. However, even though all mint C3-hydroxylases appear to require an isoleucine in position 5 after the ExxR motif, other Lamiaceae hydroxylases capable of C3-hydroxylation possess a phenylalaine in said position, just like the mint C6-hydroxylases. The *Perilla* limonene hydroxylase (CYP71D174) for example is able to perform C3-, C6-, and C7-hydroxylations (Mau et al. 2010), while the *Origanum* CYP71D178 is able to perform C2-, C3-, and C6-hydroxylations with different substrates, and the *Thymus Tv*-CYP71D180-T_{28} and *Origanum* CYP71D181 coded enzymes are able to perform C2- and C6-hydroxylations with different substrates (Crocoll 2011). These

observations indicate that, although position 5 after the ExxR motif controls part of the catalytic properties of the enzyme, it is not the only key-player and that other residues add to the specification of the hydroxylation position of the substrate. Thus different residues could be observed in position 7 and 8 after the ExxR motif but no conclusions can be inferred from these (Figure 3-7).

Another likely candidate for the control of substrate specificity and regioselectivity could be the corresponding residue of the Phe-87 position in CYP102A, the P450 BM-3 of *Bacillus megaterium*, situated in the B-C loop region (Vottero et al. 2011). This residue, located in SRS 1, is assumed to be a hotspot for regioselectivity. However, since the B-C loop region is highly variable it is difficult to predict the exact corresponding residue location in the Lamiaceae CYP71D sequences without appropriate structural information. Nonetheless predictions were made for several thousand cytochrome P450 sequences, revealing the residues at this position are predominantly of an aliphatic nature or a phenylalanine (Sirim et al. 2010). Small polar amino acids were less frequently observed and charged residues were only very infrequent. Even though no prediction to the Phe-87 corresponding residue can be made for the Lamiaceae CYP71D sequences, is it interesting to take note of the residue differences present in SRS 1 and SRS 3. In both sequence recognition sites are several amino acid residues, which could be considered as possible candidates for studies aided by site-directed mutagenesis.

In addition to this have two site-directed mutagenesis studies, conducted on members of the CYP71D subfamily, demonstrated that three positions in SRS 6 can also have considerable influence on the catalytic activity (Schuler and Rupasinghe 2011). The positions found in *Nicotiana tabacum* CYP71D20 and *Hyoscyamus muticus* CYP71D55 correlate with Pro-476, Leu-478, and Gly-480 in *M. spicata* CYP71D18 and Glu-477, Leu-479, and Gly-481 in *M.* x *piperita* CYP71D15 (Figure 3-7). All Lamiaceae CYP71D members possess a leucine and a glycine in the corresponding positions but differ in the corresponding positions of Pro-476 and Leu-479, respectively. With the exception of the *P. fruticosa* CYP71D174, possess all sequences of enzymes capable of C2- and/or C6-hydroxylation (including *Tv*-CYP71D180-T_{28}) a proline in this position, preceded by a polar threonine and in the case of *Ov* CYP71D178 a nonpolar lysine. *Pf* CYP71D174 has a lysine in the corresponding position preceded by a nonpolar alanine. Like *Pf* CYP71D174 and *Ov* CYP71D178, both capable of C3-hydroxylation, have the mint P450s also capable of C3-hydroxylation a nonpolar residue

(alanine), which is followed by an acidic polar and negatively charged glutamic acid. The cytochrome P450s *Tv*-CYP71D179-L_{48}, *Ov* CYP71D179, and *Tv*-CYP71D182, whose enzymatic activity is currently unknown, possess a threonine preceded by a lysine in said positions. It seems likely that these differences play a significant role in the selectivity and specificity of the proteins. Additional points of interest in the sequence recognition sites might be the Phe/Tyr and the Trp/Lys residue differences found in SRS 2, respectively. It is most likely that several other amino acid differences not necessarily found in the SRSs, account for the observed and predicted regio- and stereospecific deviation of the Lamiaceae CYP71D members.

4 Molecular control of the monoterpene polymorphism in *Thymus vulgaris*

Introduction

The essential oil composition of natural populations of *Thymus vulgaris*, in particular the identity of the major monoterpenes, can vary greatly between individual plants, as we have seen in chapter one. These various monoterpene chemotypes are named after the dominant compound found in each one, and seven chemotypes have been described in *T. vulgaris*: five non-phenolic monoterpene chemotypes (geraniol (G), 1,8-cineole, *α*-terpineol (A), *trans*-sabinene hydrate (U), and linalool (L)), and two phenolic monoterpene chemotypes (thymol (T) and carvacrol (C)). To study the molecular basis of these chemotypes, we isolated several genes involved in essential oil biosynthesis in thyme, as reported in chapter two. The isolated genes include two linalool synthases (*Tv*TPS2 and *Tv*TPS3), two sabinene hydrate synthases (*Tv*TPS6 and *Tv*TPS7)(Krause et al. 2013), one *γ*-terpinene synthase (*Tv*TPS1), an (*E*)-*β*-caryophyllene synthase (*Tv*TPS9), and two cytochrome P450 monooxygenases (*Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28}). These gene sequences can now be used as probes to investigate how the compositional differences in essential oil formation are controlled. This information might be applicable to other species in which monoterpene chemotypes have been reported, such as other members of the Lamiaceae plant family including mint, lavender, thyme and bergamot, as well as genera of other families such as *Artemisia* (Asteraceae), *Melaleuca* and *Eucalyptus* (Myrtaceae), or *Picea* (Pinaceae).

The low number of chemotypes reported in *T. vulgaris*, when compared to other closely related species, as for example *T. praecox* subsp. *arcticus* with as many as 17 different chemotypes (Schmidt et al. 2004), has made this plant a good candidate for the study of the genetic basis and control of chemotype formation in the genus *Thymus* in the narrower sense and the Lamiaceae in the broader sense. Accordingly, extensive controlled crosses of *T. vulgaris*, as well as sampling of natural populations were conducted in the 1980's in southern France (Vernet et al. 1986). The chemical variation was demonstrated to be under strong genetic control. Variation appears to be determined by an epistatic series of

six biosynthetic loci (G → A → U → L → C → T). Thus, a plant with the dominant G-allele will exhibit the geraniol phenotype, regardless of whether it possesses dominant or recessive alleles at the other loci. A linalool phenotype, on the other hand, will only be exhibited when the plant possesses a dominant L-allele while simultaneously being recessive in the G-, A-, and U-loci. The thymol chemotype is therefore only exhibited in plants which are homozyogous recessive at all six loci.

It is well known that the regulation of any metabolic pathway can occur at several levels. Control is mediated by factors such as, which genes are present, the rates of gene transcription and translation, and the level of enzyme activity. Several factors or combinations of factors could control monoterpene biosynthesis in *T. vulgaris*. A simple scenario would be that the absence or presence of genes encoding terpene synthases determines the terpene composition of each *T. vulgaris* chemotype. However, this possibility is very unlikely considering the pattern of epistasis observed in the previously conducted crossing experiments (Vernet et al. 1986). More likely would be differences amongst the different chemotypes in the coding region of terpene synthase genes resulting in loss of function or differences in the non-coding introns resulting in different splicing patterns. It is also possible that neither the coding nor the non-coding regions of the terpene synthase genes differ, but rather that the point of control can be found in the promoter region.

Several studies have shown that differences in terpene production observed in plants correlate with differences in the transcript levels of terpene synthase genes. The expression patterns of *Zea mays* (*E*)-*β*-caryophyllene synthase, for example, differs not only distinctly between individual plant organs but also between European maize lines and most American maize varieties (Köllner et al. 2008). Most North American maize lines, in contrast to European maize lines, do not produce (*E*)-*β*-caryophyllene in response to herbivore damage, which is also mirrored in a reduced or nonexistent transcript level of the terpene synthase gene involved - tps23. It could be shown that nonproducing American maize lines possess an active tps23 allele in their genomes, but lack transcript. Similar examples for the correlation between transcript level and terpene production in the Lamiaceae plant family include the close relationship of the geraniol concentrations in two *Ocimum basilicum* cultivars to the presence or absence of geraniol synthase transcript in these cultivars (Iijima et al. 2004b), the correlation of terpene expression patterns with terpene synthase transcripts in *Origanum vulgare* lines (Crocoll et al. 2010), as well as the correspondence of 1,8-cineole synthase

expression patterns and 1,8-cineole concentrations in *Lavandula angustifolia*, *L. latifolia* and their cross *L.* x *intermedia* (Demissie et al. 2012). It is therefore likely that part of the control of chemotype formation in *T. vulgaris* also occurs on the transcriptional level.

Material and Methods

Isolation of genomic DNA from thyme leaves

The following protocol was applied to isolate genomic DNA (gDNA) from thyme: young leaf tissue was harvested and ground in liquid nitrogen with mortar and pestle (see Chapter I). The ground material was scraped into a chilled 50 ml tube and 5 ml of CTAB isolation buffer (1.4 M NaCl, 100 mM Tris-HCl pH 8.0, 20 mM Na-EDTA, 2 % [w/v] CTAB [ethyltrimethylammonium bromide], 0.2 % [v/v] 2-mercaptoethanol) per gram tissue and preheated to 60 °C, were added. Samples were then incubated at 60 °C for 30 min with occasional gentle swirling, cooled briefly on ice and mixed with the same volume of phenol:chloroform (1:1). To prevent the precipitation of CTAB at temperatures below 15 °C, the centrifugation and working steps were performed at room temperature if not stated otherwise. After centrifugation at 16.000 x g for 10 min the aqueous phase was removed with a wide-bore pipette to a new 50 ml tube. The same volume of chloroform:isoamyl alcohol (24:1) was added and the sample was again centrifuged at 16.000 x g for 10 min. The upper aqueous phase was transferred to a new 50 ml tube and two-thirds volume of ice cold 100 % 2-propanol were added to precipitate the nucleic acids and centrifuged at 5.000 x g for 10 min. The supernatant was discarded, 10 ml wash buffer (76 % [v/v] EtOH and 10 mM ammonium acetate) were added, and incubated for a minimum of 20 min. The sample was centrifuged at 16.000 x g for 10 min at 4 °C, the supernatant discarded and the pellet was allowed to dry. The pellet was then resuspended in 3 ml TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM Na-EDTA pH 8.0) and RNase A was added to a final concentration of 10 µg ml^{-1} and incubated for 30 min at 37 °C. The DNA was extracted once with an equal volume of phenol:chloroform (1:1) and centrifuged at 16.000 x g for 10 min at 4 °C. The upper aqueous phase was collected and ammonium acetate with pH 7.7 was added to a final concentration of 2.5 M together with 2.5 volumes of ice-cold 100 % EtOH to precipitate the DNA. The DNA was centrifugated for a last time at 4 °C for 10 min at 10.000 x g, and the pelleted DNA was then washed three times with ice-cold 70 % [v/v] EtOH. The DNA was finally resuspended in 200 - 500 µl of TE buffer (pH 8.0) and stored at -20 °C.

Isolation of genomic DNA sequences of *Tv*TPS1 from different chemotypes

The gene-specific primers gDNA-*Tv*TPS1-fwd and gDNA-*Tv*TPS1-rev (see Supplementary Material, Table S1) were used for the amplification of the *Tv*TPS1 gene from the isolated genomic DNA from thyme chemotypes G_1, A_{10}, U_6, L_{48}, and T_{28}. The PCR

reaction contained 2 µL buffer, 1 µl dNTP (2.5 mM each), 0.5 µl Advantage Taq cDNA polymerase (5 U / µl), 1 µl of each gene-specific primer (10 µM each), 2 µl gDNA (4 - 17 ng / µl), 12.5 µl *a. dest*.. PCR was performed using a denaturing step at 96 °C for 5 min, followed by 33 cycles (denaturation at 96 °C for 40 sec, annealing at 60 °C for 40 sec, extension at 68 °C for 3 min), and a final extension step at 68 °C for 15 min, followed by a 4 °C hold. The PCR products were gel-purified, subcloned in pCR4-TOPO, sequenced and subsequently analysed as described in Chapter II.

Southern-blot analysis

Isolated gDNA from thyme chemotypes L_{48} and T_{28} were digested with BspHI, KpnI, SalI, SmaI, and XmaI in the buffer proposed by the manufacturer (NEB, Schwalbach, Germany, and Invitrogen, Carlsbad, USA). The reaction mixture contained 6 µg gDNA, 50 – 100 U enzyme, and 3 µl reaction buffer in a total reaction volume of 30 µl, and was incubated overnight at 37 °C. The applied reaction conditions were chosen to avoid possible non-specific activity of the enzymes. Digested gDNA (10 µg / lane) was mixed with 1/6 volume loading buffer (50 % [v/v] glycerol, 0.05 % [w/v] bromophenol blue, 0.04 % [w/v] xylene cyanol, 100 mM Na-EDTA pH 8.0) and electrophoresed in a 1.2 % agarose gel (1.2 % [w/v] agarose, 1 µg ml^{-1} EtBr, 0.5 x TAE buffer: 20 mM Tris acetate pH 8.2, 1 mM Na-EDTA) in 0.5 x TAE buffer in an EC360M Maxiwell system (EC Apparatus Corporation, St. Petersburg, USA) over night at 80 V. To compare DNA fragments' lengths, a 10 kb marker (Invitrogen) was separated together with the DNA samples. To assess equal loading the gel documentation system GeneGenius and the software GeneSnap and GeneTools from Syngene (Cambridge, UK) were used. To transfer the DNA by capillary blotting onto Hybond-XL nylon membranes (Amersham Biosciences, Uppsala, Sweden), the gel was first depurinated in 0.25 M HCl, then incubated for 40 min in denaturing solution (0.4 M NaOH, 1.5 M NaCl), before being incubated two times for 20 min in neutralization solution (0.5 M Tris-HCl pH 7.0, 1.5 M NaCl). Capillary blotting was performed overnight with 20 x SSC (3 M NaCl, 0.3 M trisodium citrate dihydrate pH 7.0) and the transferred DNA was fixed to the filter by UV-crosslinking in an UV-Stratalinker® 2400 (Stratagene, La Jolla, USA) two times with a fluence of 120 mJ $cm^{2\ -1}$. For both prehybridization and hybridization, UltraHyb™ buffer (Ambion) was used, and probes were labeled with [α-^{32}P]dATP (20 µCi) using the Strip-EZ™ PCR kit from Ambion according to the manufacturer's instructions. Probes were amplified as ssDNA from a short linearized DNA fragment of *Tv*TPS1 (bp 311) using the gene sequence from chemotype T_{28}. For primer details see Supplementary Material, Table S1.

After amplification unbound nucleotides were removed by gel filtration on ProbeQuant™ G-50 Micro columns (Amersham Biosciences, Uppsala, Sweden), and purified probes (6 x 10^6 CPM ml^{-1} hybridization buffer) were directly applied into the hybridization buffer and membranes were incubated overnight at 42 °C. The membranes were washed once with 2 x SSC at 65 °C, three times with 1 x SSC at 65 °C, once with 0.5 x SSC at 65 °C, and twice with 0.1 x SSC at 65 °C, while the radioactivity was monitored with a Geiger Müller counter (Morgan series 900 mini-monitor; Mini Instruments Ltd., Burnham-on-Crouch, UK). All SSC buffers contained 0.1 % [w/v] SDS. The washed membranes were sealed in PVC bags and exposed to BioMax MS-1 x-ray films (Eastman Kodak Company, Carestream Health, Rochester, NY, USA) with an intensifying screen over night at -80 °C. Following the manufacturer's instructions the films were developed with the Medical Film processor SRX-101 (Konica, Tokyo, Japan). In addition, storage phosphor screens (Amersham Bioscience) were exposed to the membranes for 3 – 4 h and analysed on the phosphor imaging system Storm™ 840 using the software GeneQuant and GeneScanner (all from Molecular Dynamics, Sunnyvale, CA, USA).

Northern-blot analysis

Whole leaf RNA from chemotypes G_1, A_{10}, U_6, L_{48}, C_{10}, T_2, and T_{28}, was isolated as described in Chapter II and quantified by spectrophotometry and gel electrophoresis. Total RNA (7 µg / lane) was denatured in 1 volume formaldehyde loading dye (Ambion) at 65 °C for 10 min and electrophoresed in a denaturing RNA gel (1 % [w/v] agarose, 10 % (v/v) NorthernMax™ denaturing gel buffer (Ambion), 1 µg ml^{-1} EtBr) using NorthernMax™ running buffer (Ambion) in an EC360M Maxiwell system at 80 V following the manufacturer's instructions. The visualization before blotting was performed as described above. The gel was then rinsed three times with 0.1 % [v/v] diethylpyrocarbonate-treated *a. dest.*, before being incubated in 0.05 M NaOH for 10 min and equilibrated for 20 min in 20 x SSC. The RNA was then blotted onto Hybond-XL nylon membranes (Ambersham) with 20 x SSC by capillary transfer overnight and UV-crosslinked to the membrane two times with a fluence of 120 mJ cm^{2-1} the next day. Fragments with large sequence differences from the corresponding regions of other terpene synthase and P450 genes were chosen (*Tvtps1*: 550 bp, *Tvtps2*: 420 bp, *Tvtps7*: 243 bp, *Tvtps9*: 576 bp, and *Tvp450*: 531 bp for primer details see Supplementary Material, Table S1) and amplified as described above from short fragments of the corresponding enzyme open reading frames.

Results

Regulation of *Thymus vulgaris* chemotypes at the genomic level

Analyzing the genomic sequence of the γ-terpinene synthase TvTPS1 gene in T. vulgaris

The γ-terpinene synthase gene (*Tv*TPS1) encodes a protein that produces an intermediate of thymol and carvacrol biosynthesis. Thus this gene is required in all T and C chemotypes. To determine how widespread this gene is among other *T. vulgaris* chemotypes, we isolated genomic DNA from each one and attempted to clone *Tv*TPS1. The resulting sequences were compared. Genomic DNA was isolated from young leaves of the chemotypes G_1, A_{10}, U_6, L_{48}, and T_{28}. PCR amplification with gene specific primers for *Tv*TPS1 resulted in a single amplicon for each chemotype, which was subsequently sequenced. The length of the isolated sequences of *Tv*TPS1 differed between the chemotypes, with 2270 bp, 2291 bp, 2303 bp, 2296 bp, and 2299 bp for the chemotypes G_1, A_{10}, U_6, L_{48}, and T_{28}, respectively. The genomic DNAs thus exceed the open reading frame of the *Tv*TPS1 cDNA by 489 to 522 bp.

The genomic sequences were aligned with the *Tv*TPS1 cDNA sequence by hand, and the intron/exon boundaries of the genomic sequence were determined. The genomic DNA from each of the chemotypes studied had identical exon sequences at the amino acid level, but the sequences differed in the length and composition of their individual introns (Figure 4-1). Each sequence contained six introns with a size from 65 to 115 bp and seven exons with a size from 138 to 383 bp. The observed intron-exon pattern of six introns and seven exons, is typical for TPS genes of the class III group (Trapp and Croteau 2001), much like the limonene synthases from *Perilla frutescens* (Yuba et al. 1996) and *Arabidopsis thaliana* (Trapp and Croteau 2001), and the 1,8-cineole synthases of *Lavandula latifolia*, *L. angustifolia*, and *L.* x *intermedia* (Demissie et al. 2012). The intron and exon sizes are also similar to the sizes generally observed in this class.

The intron-exon junction sites of the *Tv*TPS1 sequences followed the standard GT/AG rules for *cis*-splicing (Blumenthal and Steward 1997), and the introns contained ~60 % AU nucleotides, typically required for efficient splicing in dicotyledons. In contrast monocotyledons are more flexible and can splice introns with as little as 30 % AU (Brown and Simpson 1998). In addition, the phasing of the intron insertion sites - how codons of the open reading frame are split across intron boundaries - followed the same pattern as in other class III TPS genes and appears to be equally well conserved (Table 4-1). Intron phase

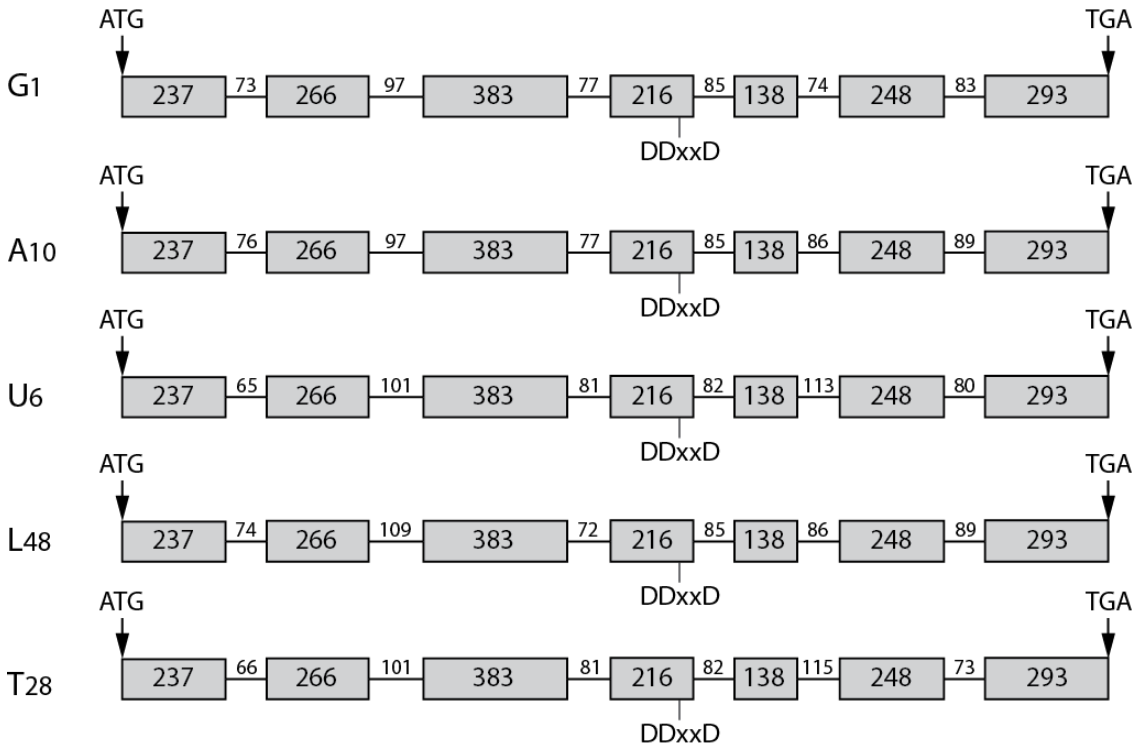


Figure 4-1: Schematic display of the intron-exon structure of the *Tv*TPS1 gene of different chemotypes. Exons are represented by grey boxes, and introns are represented by lines connecting adjacent exons. The length of nucleotide sequences is shown for exons in the boxes and for introns above the lines. The start and stop codons are indicated and DDxxD marks the position of the conserved aspartate-rich region.

depends on whether the intron is located before the first, after the first, or after the second nucleotide of a codon, phase '0', '1' or '2', respectively (Long et al. 1995). The introns III, VIII, XI, XII, XIII, and XIV show the phases '0', '1', '2', '2', '0', and '0', respectively (Table 4-1).

Table 4-1: Phases of introns within the *T. vulgaris* γ-terpinene synthase and in comparison with those of basal angiosperm, angiosperm, and gymnosperm mono- and sesquiterpene synthases. Roman numerals represent intron numbers. Intron phase numbers '0', '1', and '2' refer to the intron insertion before the first nucleotide, after the first nucleotide or after the second nucleotide of a codon, respectively. Dashes indicate the absence of an intron. M11 (*Mg*) refers to putative monoterpene synthase Mg11 from *Magnolia grandiflora*; S17 (*Mg*) refers to mono/sesquiterpene synthase Mg17 from *M. grandiflora*; LS (*Pf*) refers to limonene synthase from *Perilla frutescens*; CS (*L*) refers to 1,8-cineole synthases from *Lavandula angustifolia*, *L. latifolia*, and *L.* x *intermedia*; γTS (*Tv*) refers to γ-terpinene synthase from *Thymus vulgaris*; γTS (*Tc*) refers to γ-terpinene synthase from *T. caespititius*;

*α*TS (*Tc*) refers to *α*-terpineol synthase from *Thymus caespititius*; CS (Ga) refers to *δ*-cadinene synthase from *Gossypium arboreum*; ES (*Nt*) refers to 5-epi-aristolochene synthase from *Nicotiana tabacum*; LS (*Ag*) refers to limonene synthase from *Abies grandis*; SS (*Ag*) refers to selinene synthase from *A. grandis*; PS (*Ag*) refers to (-)-pinene synthase from *A. grandis*; BS (*Ag*) refers to *α*-bisabolene synthase from *A. grandis*; TS (*Tb*) refers to taxadiene synthase from *Taxus brevifolia*; and AS (*Ag*) refers to abietadiene synthase from *A. grandis*. All data except for that from the *T. vulgaris γ*-terpinene synthase were adapted from literature (Lee and Chappell 2008, Demissie et al. 2012, Lima et al. 2013).

Type	Intron	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Basal Angiosperms	M11 (*Mg*)	-	-	0	-	-	-	-	1	-	-	2	2	-	0
	S17 (*Mg*)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
Angiosperms	LS (*Pf*)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
	CS (*L*)	-	-	0	-	-	-	-	0	-	-	1	1	0	0
	γTS (*Tv*)	**-**	**-**	**0**	**-**	**-**	**-**	**-**	**1**	**-**	**-**	**2**	**2**	**0**	**0**
	*γ*TS (Tc)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
	*α*TS (Tc)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
	CS (*Ga*)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
	ES (*Nt*)	-	-	0	-	-	-	-	1	-	-	2	2	0	0
Gymnosperms	LS (*Ag*)	-	-	0	-	-	-	2	1	0	2	2	2	0	0
	SS (*Ag*)	-	-	0	-	-	-	2	1	0	2	2	2	0	0
	PS (*Ag*)	-	-	0	-	-	-	2	1	0	2	2	2	0	0
	BS (*Ag*)	-	-	0	1	2	1	2	1	0	2	2	2	0	-
	TS (*Tb*)	-	-	0	1	2	1	2	1	1	2	2	2	0	0
	AS (*Ag*)	1	1	0	1	2	1	2	1	0	2	2	2	0	1

Southern blot analysis of *Tv*TPS1 gene

To compare the *Tv*TPS1 gene copy number in various chemotypes, Southern blots were carried out on the geraniol chemotype G_1, the linalool chemotype L_{48}, and the carvacrol/thymol-hybrid chemotype T_{28} using five restriction enzymes (BspHI, EcoRI, HindIII, XhoI, and XmaI). These enzymes were selected because their restriction sites were not found in the probe sequence used. An [α-^{32}P]dATP-labelled PCR product (311 bp), which did not include introns, was used as the probe. We found three distinct bands in all lanes of the L_{48}- and T_{28}-chemotypes, while the G_1-chemotype lanes had between five to six bands (Figure 4-2). This indicated that between three and six copies of the *γ*-terpinene synthase *Tv*TPS1 gene exist in the three tested *T. vulgaris* chemotypes.

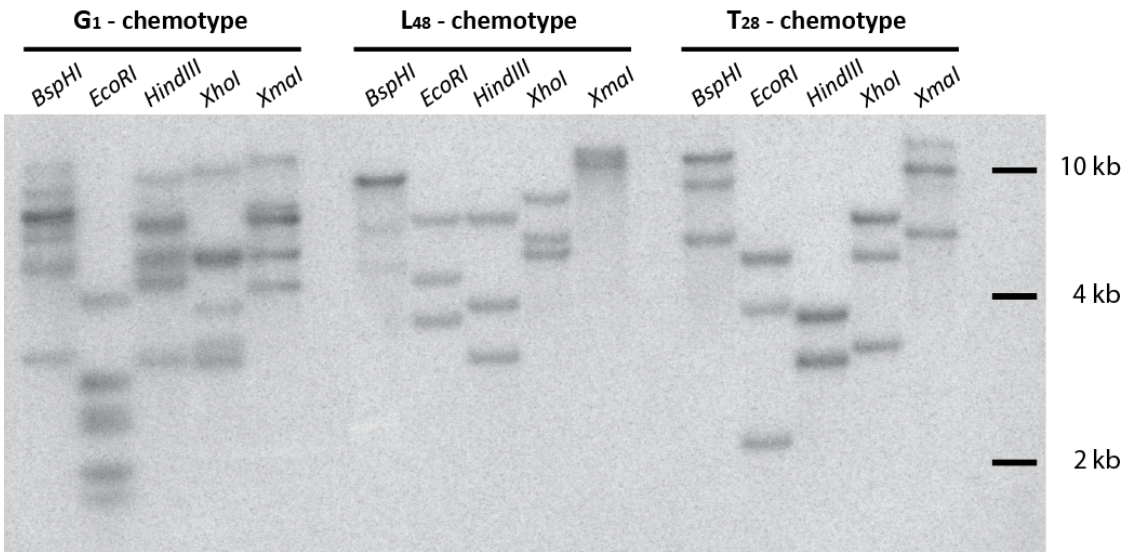


Figure 4-2: Determination of genomic copy number of the *Tv*TPS1 gene in selected *T. vulgaris* chemotypes. Southern blot-based estimation of *Tv*TPS1 gene copy numbers in *T. vulgaris* chemotypes G_1, L_{48}, and T_{28}. Genomic DNA (6 µg per lane) was digested with the indicated restriction enzymes and hybridized against a 311 bp, *Tv*TPS1-specific, [α-^{32}P]dATP-labled probe amplified from *Tv*TPS1 cDNA (see Supplementary Material, Table S1). The positions of DNA size markers (kb) are shown.

Regulation of *T. vulgaris* chemotypes at the transcript level

The transcript levels of the genes described in Chapter II were also examined in each chemotype via RNA gel blotting to determine how these correlated with the essential oil composition. RNA samples were extracted from young expanding leaves of all seven *T. vulgaris* chemotypes investigated in this thesis (G_1 – geraniol chemotype; A_{10} – *α*-terpineol chemotype; U_6 – *trans*-sabinene hydrate chemotype; L_{48} – linalool synthase; C_{10} – carvacrol chemotype; T_2 – carvacrol/thymol chemotype; T_{28} – carvacrol/thymol chemotype) and analyzed (Figure 4-3 to Figure 4-6).

TvTPS1 and the two putative cytochrome P450s

Figure 4-3 shows the signals observed in the different chemotypes with a probe (550 bp) designed for the *γ*-terpinene synthase *Tv*TPS1 isolated from the T_{28}-chemotype, and a probe (531 bp) designed for the *Tv*-CYP71D180-T_{28} and *Tv*-CYP71D179-L_{48} sequences, which are 82 % identical, and amplified from the *Tv*-CYP71D180-T_{28} template. Both probes were hybridized at 42 °C overnight and the blot then washed under high stringency conditions. A strong *Tv*TPS1 signal can be observed in the four chemotypes L_{48}, C_{10}, T_2, and

T_{28} while only faint signals are perceived in the chemotypes A_{10} and U_6. A similar signal distribution was produced with the *Tv*-CYP71D180-T_{28} probe with no signals observed in the chemotypes G_1, A_{10}, and U_6. The strongest signal is seen with T_{28}-RNA followed by C_{10}-RNA and L_{48}-RNA, while T_2-RNA only produced a faint signal.

When the results of the RNA hybridization are compared with the terpene content of the thyme plant chemotypes, a correlation between transcript abundance and monoterpene content becomes evident (Figure 4-3). *γ*-Terpinene, the main product of *Tv*TPS1, and *α*-terpinene, the principal minor product of *Tv*TPS1, can be found in the essential oil of all chemotypes in which this gene transcript is present, and the strength of signal seems to correlate strongly with the amount of *γ*-terpinene and *α*-terpinene found in the essential oil. Correlation of this transcript is also strong with the amounts of thymol, carvacrol and *p*-cymene, all thought to be oxidized metabolites of *γ*-terpinene. The metabolites thymol, carvacrol and *p*-cymene can also be found in the essential oil of all chemotypes which have transcripts for the putative cytochrome P450 gene *Tv*-CYP71D180-T_{28}.

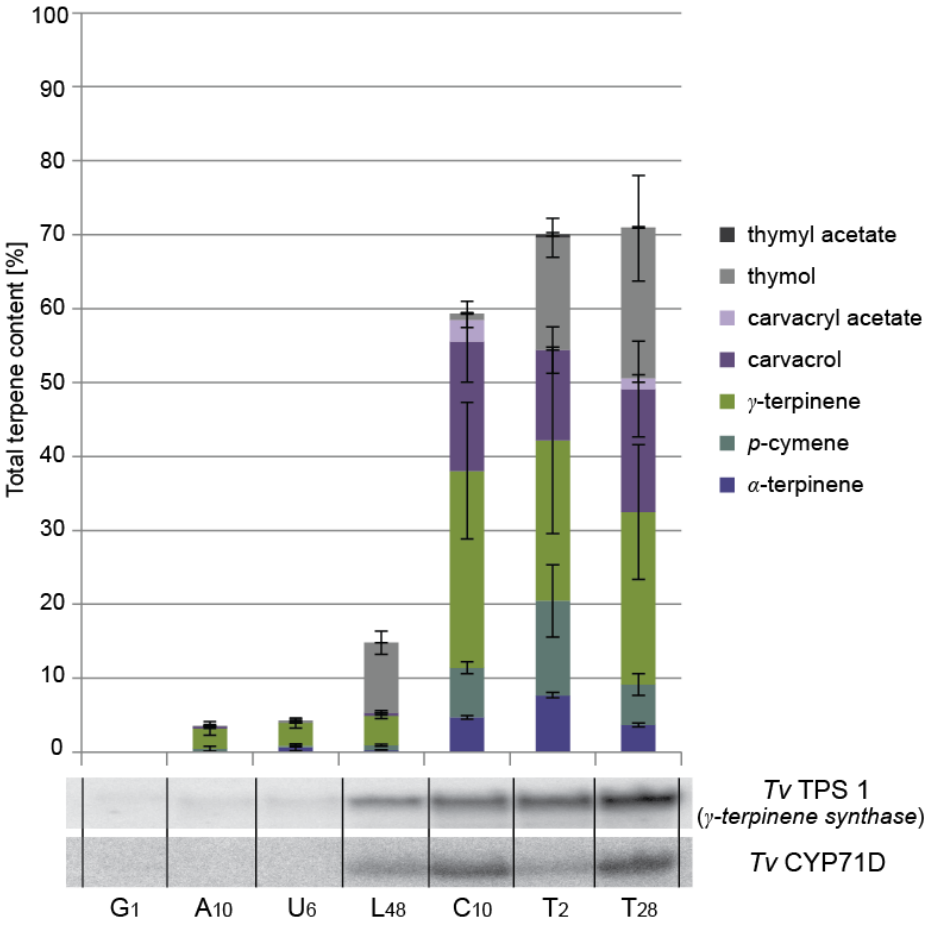


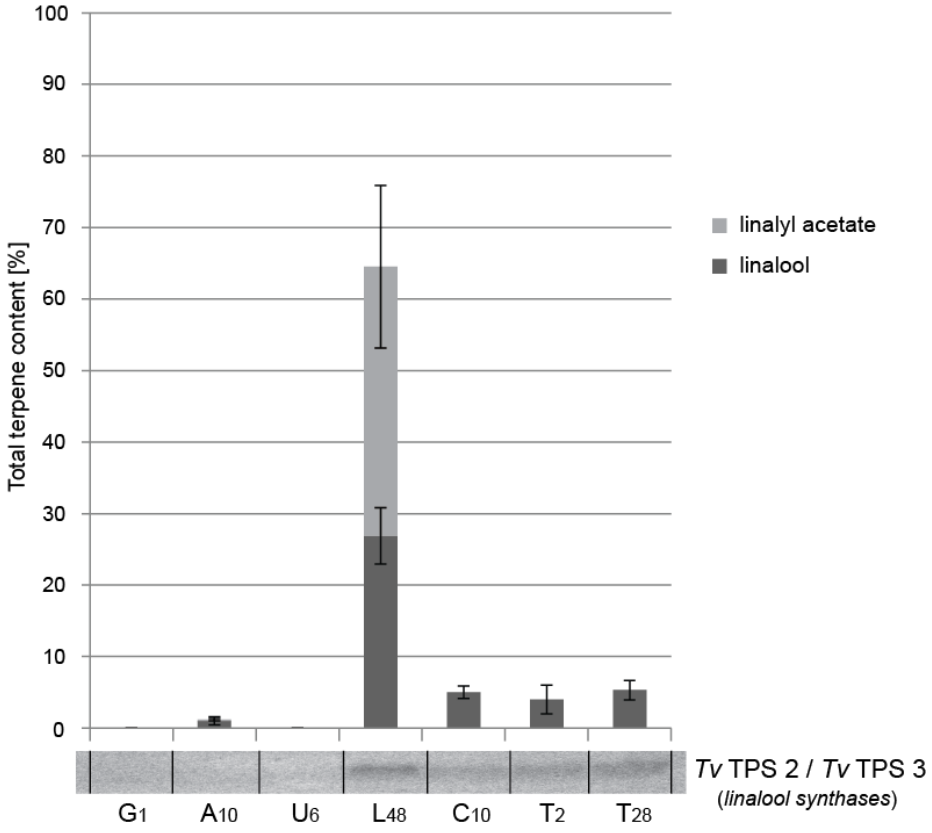
Figure 4-3: Expression analysis of *Tv*TPS1 and the *Tv*-CYP71D180-T_{28} genes *in planta* compared to essential oil content of different *T. vulgaris* chemotypes. The amounts of *α*-terpinene, *p*-cymene, *γ*-terpinene, carvacrol, carvacryl acetate, thymol, and thymyl acetate, found in the terpene profiles of the seven *T. vulgaris* chemotypes analyzed in this thesis are shown as stacked columns. *γ*-Terpinene is the main product and *α*-terpinene is the principal side product of *Tv*TPS1, while the other compounds are thought to be formed via a putative *Tv*-CYP71D monooxygenase. Mono- and sesquiterpenes were extracted, identified and quantified by GC-MS. Each bar represents the mean values ± SE of three biological and three technical replicates, except for the G_1-chemotype, where two instead of three technical replicates were used. The panels show RNA hybridization analyses (7 µg/lane) for the transcript levels of the *γ*-terpinene synthase *Tv*TPS1 and the two CYP71D candidates (*Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28}) in leaves of the seven *T. vulgaris* chemotypes analyzed in this thesis. G_1 – geraniol chemotype; A_{10} – *α*-terpineol chemotype; U_6 – *trans*-sabinene hydrate chemotype; L_{48} – linalool synthase; C_{10} – carvacrol chemotype; T_2 – carvacrol/thymol chemotype; T_{28} – carvacrol/thymol chemotype.

TvTPS2 and TvTPS3

To analyze the transcript levels of the two linalool synthases *Tv*TPS2 and *Tv*TPS3 which are 91 % identical to each other on the nucleotide level, a 420 bp long probe was designed using the *Tv*TPS2 sequence as template. The probe was hybridized over night at 42 °C and the blot was washed under highly stringent conditions. There was an excellent correlation between the *Tv*TPS2/TPS3 probe and linalool content. A comparatively strong probe signal was observed for the linalool chemotype L_{48}, which has a high content of linalool, while lesser signals were visible for the phenolic chemotypes C_{10}, T_2, and T_{28} which have low amounts of linalool (Figure 4-4). No probe hybridized in the lanes of the other three chemotypes, and these had no or only traces of linalool in their essential oil. The probe signals seem to strongly correlate with the linalool concentration found in the essential oil of the tested chemotypes, with the exception of the *α*-terpineol chemotype A_{10}, which possesses about 1 % linalool in its essential oils while showing no transcript signal for the linalool synthases tested.

Figure 4-4: Expression analysis of *Tv*TPS2 and *Tv*TPS3 *in planta* compared to the essential oil content of the different *T. vulgaris* chemotypes. The amounts of linalool and linalyl acetate, found in the terpene profiles of the seven *T. vulgaris* chemotypes analyzed in this thesis are shown as stacked columns. Mono- and sesquiterpenes were extracted, identified and quantified by GC-MS. Each bar represents the mean values ± SE of three biological and three technical replicates, except for the G_1-chemotype, where two instead of three technical replicates were used. The panel shows the RNA

hybridization analysis (7 µg/lane) for the transcript levels of the two linalool synthases *Tv*TPS2 and *Tv*TPS3 in leaves of the seven *T. vulgaris* chemotypes analyzed in this thesis. G_1 – geraniol chemotype; A_{10} – *α*-terpineol chemotype; U_6 – *trans*-sabinene hydrate chemotype; L_{48} – linalool synthase; C_{10} – carvacrol chemotype; T_2 – carvacrol/thymol chemotype; T_{28} – carvacrol/thymol chemotype.



TvTPS6 and TvTPS7

The transcript levels of the two sabinene hydrate synthases *Tv*TPS6 and *Tv*TPS7 were compared to the amounts of the sabinene hydrates and limonene, a principal minor product of both synthases (Figure 4-5). The sequences of *Tv*TPS6, a (*Z*)-sabinene hydrate synthase, and *Tv*TPS7, an (*E*)-sabinene hydrate synthase, are 89 % identical on the nucleotide level and both encode multiproduct synthases, differing mainly in the amount of (*Z*)- and (*E*)-sabinene hydrate produced (Krause et al. 2013). The probe, which was designed with *Tv*TPS7 as a template, was 243 bp long and 89 % identical (26 / 243 nucleotide differences) to the sequence of *Tv*TPS6. After an overnight hybridization at 42 °C the blot was washed under highly stringent conditions. While the overall signal is not very strong it is obvious that both genes seem to be only expressed in significant amounts in two chemotypes, the sabinene-

hydrate chemotype U_6 and the α-terpineol chemotype A_{10}. In all other chemotypes no probe signal could be perceived. The amount and composition of monoterpenes found in the U_6-chemotype ((*E*)- and (*Z*)-sabinene hydrate and limonene) correlate well with the transcript abundance of *Tv*TPS6 and *Tv*TPS7 (encoding enzymes forming (*E*)- and (*Z*)-sabinene hydrate and limonene) in this chemotype. However, the transcript of *Tv*TPS6/TPS7 in A_{10} correlates with limonene (3.02 %), but not with the low amounts of (*E*)-sabinene hydrate (0.26 %). In addition, (*E*)-sabinene hydrate (2.64 – 3.63 %) and limonene (0.94 – 1.43 %) were both present in the essential oil of the phenolic monoterpene chemotypes C_{10}, T_2, and T_{28}, but no probe signal was obtained in the RNA-hybridization analysis.

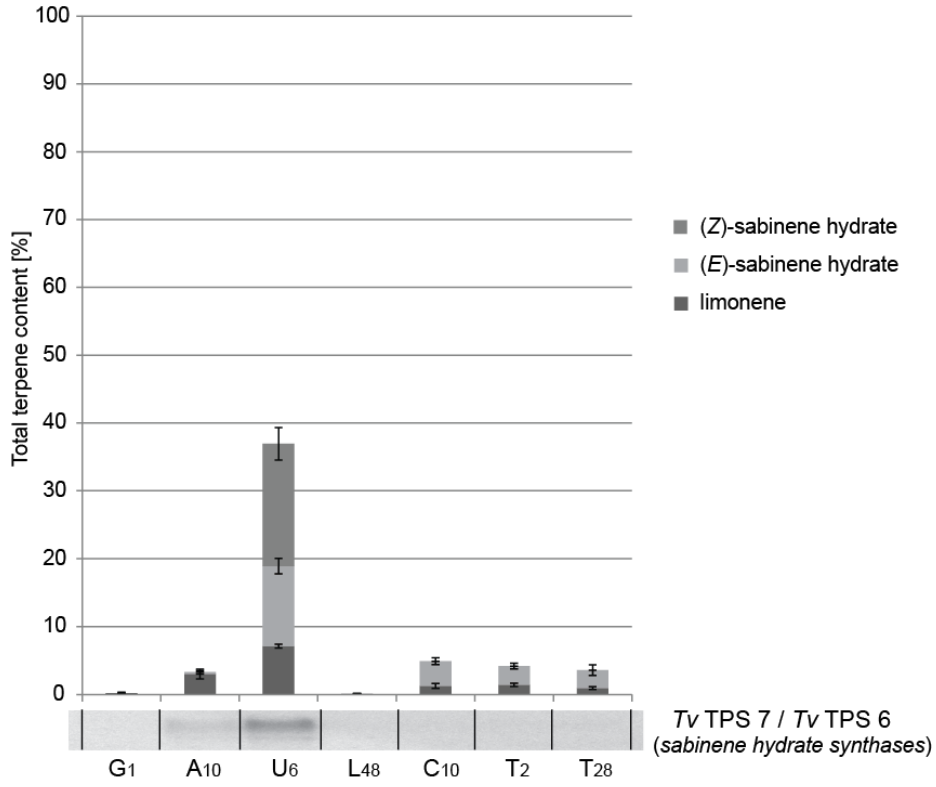


Figure 4-5: Expression analysis of *Tv*TPS6 and *Tv*TPS7 *in planta* compared to essential oil content of different *T. vulgaris* chemotypes. The amounts of the main products (*E*)- and (*Z*)-sabinene hydrate and the main side product, limonene found in the terpene profiles of the seven *T. vulgaris* chemotypes analyzed in this thesis are shown as stacked columns. Mono- and sesquiterpenes were extracted, identified and quantified by GC-MS. Each bar represents the mean values ± SE of three biological and three technical replicates, except for the G_1-chemotype, where two instead of three technical replicates were used. The panel shows the RNA hybridization analysis (6 µg/lane) for the transcript levels of the (*Z*)-sabinene hydrate synthase *Tv*TPS6 and the (*E*)-sabinene hydrate synthase

*Tv*TPS7 in leaves of the seven *T. vulgaris* chemotypes analyzed in this thesis. G_1 – geraniol chemotype; A_{10} – *α*-terpineol chemotype; U_6 – *trans*-sabinene hydrate chemotype; L_{48} – linalool synthase; C_{10} – carvacrol chemotype; T_2 – carvacrol/thymol chemotype; T_{28} – carvacrol/thymol chemotype.

TvTPS9

The transcript abundance of the sesquiterpene synthase *Tv*TPS9, an (*E*)-*β*-caryophyllene synthase, was also examined by RNA-hybridization analysis (Figure 4-6). The 576 bp long probe was designed using the *Tv*TPS9 sequence as a template and hybridized overnight at 42 °C before the blot was washed using stringent conditions. In all examined chemotypes a signal, of about similar strength can be observed, which correlates well with the amounts of (*E*)-*β*-caryophyllene found in the essential oil.

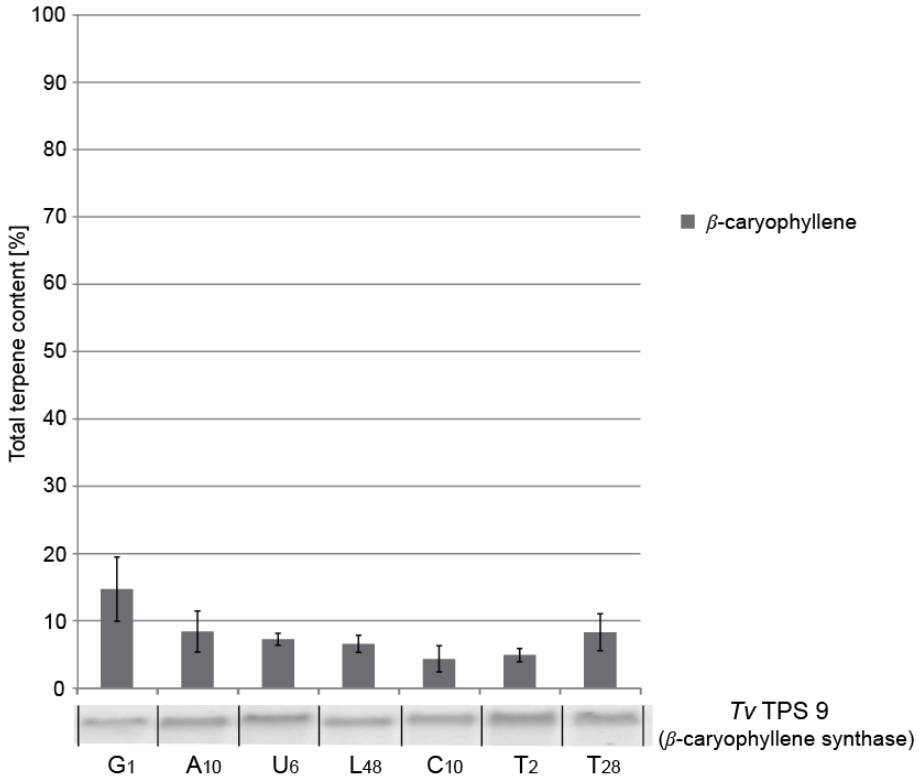


Figure 4-6: Expression analysis of *Tv*TPS9 *in planta* compared to essential oil content of different *T. vulgaris* chemotypes. The amounts of (*E*)-*β*-caryophyllene, the *in vitro* product of *Tv*TPS9, found in the terpene profiles of the seven *T. vulgaris* chemotypes analyzed in this thesis are shown as columns. Mono- and sesquiterpenes were extracted, identified and quantified by GC-MS. Each bar represents the mean values ± SE of three biological and three technical replicates, except for the G_1-chemotype, where two instead of three technical replicates were used. The panel shows the RNA

hybridization analysis (7 µg/lane) for the transcript level of the *β*-caryophyllene synthase *Tv*TPS9 in leaves of the seven *T. vulgaris* chemotypes analyzed in this thesis. G_1 – geraniol chemotype; A_{10} – *α*-terpineol chemotype; U_6 – *trans*-sabinene hydrate chemotype; L_{48} – linalool synthase; C_{10} – carvacrol chemotype; T_2 – carvacrol/thymol chemotype; T_{28} – carvacrol/thymol chemotype.

Discussion

The chemical polymorphism which affects the essential oil composition of *T. vulgaris* plants has been subject of several studies since it was first described in the 1970's (Passet 1971). Next to environmental and ecological questions the genetic heredity of chemotype formation has been the main focus. The sampling of natural populations in southern France, as well as extensive controlled crosses suggested an epistatic control underlying chemotype formation in *T. vulgaris* (Vernet et al. 1986). The crossing experiment performed by Vernet et al. (1986) led to the proposal that the G chemotype (geraniol dominant) is controlled by two different loci, while the A (*α*-terpineol dominant), U (*trans*-sabinene hydrate and others dominant), L (linalool dominant) and C (carvacrol dominant) chemotypes are controlled by a single locus each. The chemotypes can be ordered in a dominance hierarchy as follows: G → A → U → L → C → T. This means that a plant with the dominant G-allele will exhibit the geraniol phenotype, regardless of whether it possesses dominant or recessive alleles at the other loci. A linalool phenotype, on the other hand, will only be exhibited when the plant possesses a dominant L-allele while simultaneously being recessive in the G-, A-, and U-loci. The thymol phenotype, however, will only be shown when all loci possess recessive alleles.

Regulation at the level of genomic DNA

The results of these earlier studies provide a good starting point to analyze the genetic control mechanisms involved in the terpene production in the different thyme chemotypes. The presence of an epistatic series makes it unlikely for example that a 'loss of function mutation', such as the absence of a certain monoterpene synthase in a particular chemotype, is involved in the genetic control of the chemotype formation. A good candidate to investigate the genetic basis of thyme chemotypes appeared to be the *γ*-terpinene synthase *Tv*TPS1 which has been described in Chapter II. This monoterpene synthase is assumed to be a key player in the formation of the phenolic chemotypes since *γ*-terpinene is the precursor of thymol and carvacrol. But, the *γ*-terpinene synthase gene is unnecessary for the production of the major monoterpenes of the other chemotypes and so may not be present in their genomes. To test this proposition, we sequenced single amplicons of *Tv*TPS1 in the genomic DNA from the G_1, A_{10}, U_6, L_{48}, and T_{28} chemotypes, with gene specific primers and PCR (Figure 4-1). All chemotypes seemed to possess the gene. The ORFs of the genomic *Tv*TPS1 sequence, made up of seven exons, had exactly the same sequence at the amino acid level in all tested

chemotypes, and only differed slightly in length and composition of the six introns. These results allow the inference that the regulation of γ-terpinene production in thyme is not controlled by the presence or absence of a complete *Tv*TPS1 sequence from the genome of different chemotypes. The high sequence similarity among the *TvTPS1* genes in the chemotypes studies also indicates that chemotype control is not mediated by the presence of a 'loss of function' mutation in the gene or by the use of rare codons.

Regulation at the level of RNA splicing

RNAs play a central role in gene regulation at the transcriptional, post-transcriptional, and translational level in eukaryotes (Reddy et al. 2012). For example, alternative splicing (AS) is an important regulation mechanism involved in various stages of development, differentiation, and diseases in metazoan systems. Although less is known about AS in plants, it is becoming more and more evident that this mechanism fulfills important functions in plants too (Baek et al. 2008, Reddy et al. 2012). AS can take several forms, such as intron retention, exon skipping, or alternative splice site usage (e.g. tandem acceptors that result in gain or loss of three nucleotides in the spliced mRNA), and is common in land plants and animals. The frequency of splice variants differs between plants and animals, however, with intron retention being more prevalent in plants while exon skipping is predominant in animals (Baek et al. 2008, Reddy et al. 2012).

After RNA is transcribed, pre-mRNA splicing is carried out by the spliceosome, a large ribonucleoprotein complex consisting of small nuclear RNA and small nuclear ribonucleic particles. Four short core sequence elements at the exon/intron- and intron/exon-boundaries are necessary, although not sufficient, for the splice site recognition by the spliceosome. Two motifs, one at the 5'-splice site (splice donor site) with a conserved 'GU' dinucleotide and the other at the 3'-splice site (splice acceptor site) with a conserved 'AG' dinucleotide have been recognized in Metazoa. Also, a stretch of pyrimidines (polypyrimidine tract) upstream of the 3'-splice site and a branch point 17 – 40 nucleotides upstream of the polypyrimidine tract are critical. The components of the spliceosome recognize these core signals via base-pairing (Baek et al. 2008, Reddy et al. 2012). In addition to these core elements, splicing regulatory elements (SRE) are needed that function as either splicing enhancers or suppressors and can be found in either exons or introns.

To investigate the possible role that AS could play in the regulation of terpene concentrations in the thyme chemotypes, the isolated *Tv*TPS1 gDNA sequences (Figure 4-1)

were examined. The sequences are composed of seven exons and six introns and are therefore placed, like other monoterpene synthases such as the limonene synthases from *Perilla frutescens* (Yuba et al. 1996) and *Arabidopsis thaliana* (Trapp and Croteau 2001), or the 1,8-cineole synthases of *Lavandula latifolia*, *L. angustifolia*, and *L.* x *intermedia* (Demissie et al. 2012), in the Class III TPS clade as described by Trapp and Croteau (2001). The exon-intron junction sites, as have been described for *cis*-splicing, are intact and follow the GT-AG rule for splice donor and splice acceptor site. The unusual rich amount of 'AU' or 'U' typical for plant introns, relative to other eukaryotic systems, can also be found in the thyme sequences isolated. It is generally believed that a minimum of 60 % 'AU' is necessary in dicotyledonous introns for the spliceosome to function properly, while monocotyledonous introns are correctly spliced with as little as 30 % (Brown and Simpson 1998). Splicing factors and other nuclear proteins, such as plant U-rich RNA binding proteins (UBP), which have been shown to enhance intron splicing, are known to interact with the 'U'-rich pre-mRNA (Baek et al. 2008, Reddy et al. 2012). The 'AU' concentration in all six introns is above 60 percent, ranging from 64 to 88 % making it therefore unlikely that inefficient splicing is involved in the epistatic control of the monoterpene biosynthesis in thyme. However, inefficient splicing may be involved in the regulation of terpene formation in other species (Keszei et al. 2010, Lukas et al. 2010). Keszei et al. (2010) found several sequence length polymorphisms in TPS sequences isolated from *Eucalyptus rubida*, *E. sideroxylon*, *E. nicholii*, and *Callistemon citrinus*, all of which are members of the Myrtaceae. They observed inefficient RNA splicing which strongly coincided with the splicing site of exon 4, despite the fact that the sequences of surrounding splicing sites were quite different. Lukas et al. (2010) demonstrated in four *Origanum* species accessions that the presence or absence of specific patterns in intron 3 correlated with the presence or absence of certain monoterpenes in the essential oils of each accession.

Although the phasing of introns has not been shown to be involved in the splicing of pre-mRNA, it should be noted that the intron phases observed for *Tv*TPS1 are similar to the phases found in angio- and gymnosperm terpene synthase sequences (Table 4-1). Consistent with the definition of Aubourg et al. (2002) the order of intron phases in all tested chemotypes is '0' for the first intron, '1' for the second, '2' for the third, '2' for the fourth, '0' for the fifth, and '0' for the sixth intron. The 1,8-cineole synthases from the *Lavandula* species, *L. latifolia*, *L. angustifolia*, and *L.* x *intermedia* (Demissie et al. 2012), are currently the sole known exceptions to this general characteristic of class III TPS.

Regulation by gene copy number

To determine if multiple copies of the *γ*-terpinene synthase gene are present in *T. vulgaris* and if their number differs significantly between chemotypes, genomic DNA was isolated from three chemotypes spanning the whole spectrum of the epistatic series and a Southern blot analysis was performed (Figure 4-2). Chosen were the geraniol chemotype G_1, whose essential oil contains no significant amounts of *γ*-terpinene or its derivatives, the linalool chemotype L_{48}, whose essential oil contains significant amounts of *γ*-terpinene and its derivatives next to its dominant monoterpene linalool, and the carvacrol/thymol chemotype T_{28}, whose essential oil contains predominantly *γ*-terpinene and its derivatives. The gDNA was digested with 6cutter restriction enzymes (BspHI, EcoRI, HindIII, XhoI, and XmaI) which did not cut within the probe sequence of *Tv*TPS1. All of the tested *T. vulgaris* chemotypes appear to have multiple copies of the *γ*-terpinene synthase gene, with the geraniol chemotype having between five and six copies, while the linalool and carvacrol/thymol chemotypes both appear to contain three copies. These results agree with those above from the investigation of genomic DNA that the *γ*-terpinene synthase gene is present in all chemotypes. From this it can be inferred that all TPS genes in *T. vulgaris* are likely to be represented in the genome of all chemotypes regardless of the composition of the essential oil, and thus chemotype differences must be manifested at lower levels of control.

Regulation by mRNA transcript level

Since we found that the gene encoding the *γ*-terpinene synthase *Tv*TPS1 was present in all tested chemotypes, and was not subject to AS nor showed mutations in the exon sequences which would lead to premature stop codons, we examined the transcription of several TPS sequences, described in Chapter II, in comparison to the accumulation of products from their encoded enzymes (data from Chapter I). Total RNA samples, extracted from young expanding leaves of all seven chemotypes investigated in this thesis (G_1 – geraniol chemotype; A_{10} – *α*-terpineol chemotype; U_6 – *trans*-sabinene hydrate chemotype; L_{48} – linalool synthase; C_{10} – carvacrol chemotype; T_2 – carvacrol/thymol chemotype; T_{28} – carvacrol/thymol chemotype), were analyzed via RNA gel blotting and these data compared to the amount of their product(s) (Figure 4-3 to Figure 4-6). The results show a strong correlation between transcript abundance and terpene content. Although it has been previously shown that the production of certain monoterpenes is regulated through transcriptional control of the corresponding genes, this is - to the best of our knowledge - the first time that this could be proven in the context of chemotype formation. Examples of direct

correlation between monoterpene production and abundance of monoterpene synthase mRNA in Lamiaceae include studies on production of menthofuran (Mahmoud and Croteau 2003) and other monoterpenes in peppermint (Gershenzon et al. 2000, McConkey et al. 2000, Turner 2000, Turner et al. 2000), geraniol production in sweet basil (Iijima et al. 2004b), and production of various monoterpenes in oregano (Crocoll 2011). Also this correlation has been demonstrated in different *Lavandula* species (Boeckelmann 2008, Lane et al. 2010, Demissie et al. 2012). However these results are not confined to the Lamiaceae plant family but have also been reported in other plant families, such as the Scrophulariaceae (snapdragon) (Dudareva et al. 2005, Nagegowda 2010), the Myrtaceae (eucalyptus) (Padovan et al. 2013), and the Rutaceae (citrus) (Voo et al. 2012).

In *O. vulgare*, it was recently confirmed that *γ*-terpinene is the precursor of the phenolic terpenes carvacrol, thymol, and *p*-cymene, and that P450s are involved in the biosynthesis of *γ*-terpinene derivatives (Crocoll 2011), as had been proposed by Poulose and Croteau in the 1970's (1978b, a). The transcript of the *O. vulgare γ*-terpinene synthase and the P450s CYP71D180 and CYP71D182 were correlated with the amounts of *γ*-terpinene and its derivatives found in the essential oil. A similar pattern was found in the present study involving transcripts of the *γ*-terpinene synthase *Tv*TPS1 and the P450s *Tv*-CYP71D180-T_{28} and *Tv*-CTP71D182-L_{48} (Figure 4-3). The main (*γ*-terpinene) and principal minor (*α*-terpinene) products of *Tv*TPS1 were only found in the essential oil of chemotypes in which this transcript was present, and the strength of the signal correlated strongly with the amount of *γ*-terpinene and *α*-terpinene found in the essential oil. Similar results were observed for the amounts of thymol, carvacrol, and *p*-cymene, which are thought to be the oxidized metabolites of *γ* -terpinene, and were only found in the essential oil of chemotype that displayed transcript of the putative cytochrome P450 genes *Tv*-CYP71D180-T_{28} or Tv-CYP71D182-L_{48} (Figure 4-3). The results for the two linalool synthases *Tv*TPS2 and *Tv*TPS3, as well as the two sabinene hydrate synthases *Tv*TPS6 and *Tv*TPS7, in regards to mRNA transcript abundance and monoterpene content of the essential oil, are also in agreement with this trend (Figure 4-4 and Figure 4-5).

However, a few exceptions were noted to the strict correlation between terpene synthase transcript levels and product abundance in the *T. vulgaris* plants in this study. The *α*-terpineol chemotype A_{10} possesses about 1 % linalool in its essential oil but no transcript signal for the tested linalool synthases was detected in its mRNA. Also, small amounts of (*E*)-

sabinene hydrate (2.64 – 3.63 %) and limonene (0.94 – 1.43 %) were observed in the phenolic chemotypes C_{10}, T_2, and T_{28}, without the presence of any detectable transcript. The most likely explanation for this lack of congruence is that these monoterpenes are not actually products of the transcribed gene, but main or side products of currently unknown terpene synthases. Or, the detection level of the Northern blot analysis may have been so low that transcripts of the linalool and sabinene hydrate synthases were not satisfactorily visualized. In this case, both linalool and limonene are most likely side products since they can be often found in the product spectrum of multiproduct monoterpene synthases (Degenhardt et al. 2009). Sabinene hydrate is on the other hand probably a main product and its transcript would most likely have been detected if a more sensitive assay would have been used. The results do not contradict the chemical differences among chemotypes since it has been shown that these divisions do not seem to be as strict as previously thought (Thompson et al. 2003). Consistent with this generalization is the presence of linalool in the phenolic chemotypes as well as the existence of the carvacrol/thymol chemotype T_{28}.

Even though sesquiterpene biosynthesis does not appear to be controlled by the epistatic series of monoterpene chemotype formation in *Thymus vulgaris*, it can be noted that the amount of (*E*)-*β*-caryophyllene found in the essential oil correlates strongly with the transcript of the (*E*)-*β*-caryophyllene synthase *Tv*TPS9 (Figure 4-6).

Regulation by DNA methylation and other epigenetic mechanisms

In taking a complete modern view of gene regulation, we should also consider the possibility that the genetic differences among the *T. vulgaris* monoterpene chemotypes are due to epigenetic phenomena. An epigenetic trait is a stably heritable phenotype which results from changes in a chromosome without alterations in the DNA sequence itself. These traits can be maintained by DNA methylation, histone modifications (SUMOylation with Small Ubiquitin-like Modifier proteins, acetylation, methylation, and phosphorylation), histone variants, or nucleosome positioning, among other factors, thus transmitting epigenetic information from mother to daughter cell and from plant generation to plant generation (Berger et al. 2009, Zaina et al. 2010). However, genomic variation in either coding or non-coding regions of the genome may occasionally have significant influence on an epigenetic trait. For instance, single nucleotide polymorphisms found in non-coding regions can be viewed as potential phenotype modifiers rather than exclusively as unimportant and functionally neutral genetic variants. SNPs in introns, for example, have been shown to have

a significant impact on type 2 diabetes and obesity in humans (Zaina et al. 2010). These genetic variants can influence epigenetic markers, such as DNA methylation or histone markers, which can in turn provoke a rearrangement of the chromatin architecture that may have long-range effects on transcription via promoter or transcription factor binding sites (Zaina et al. 2010, Ingvarsson and Street 2011).

Examination of the genomic sequence of the *γ*-terpinene synthase *Tv*TPS1 showed that the encoded protein sequence did not differ between the tested chemotypes (Figure 4-1). However, the nucleotide sequence differed in the coding as well as in the non-coding regions. It is therefore possible that instead of alternative splicing, which we ruled out earlier, changes in the nucleotide sequence of the gene itself could have an important influence on the chromatin architecture and thus directly on the gene expression of *Tv*TPS1. To be able to make more precise predictions involving the SNPs found in the genomic sequences of *Tv*TPS1 it would be necessary to conduct further experiments with more suitable plant material, such as plants with an identical genetic background differing only in SNPs.

DNA methylation commonly occurs in plants at cytosine bases in the sequences ‘GC’, ‘CHG’, and ‘CHH’ (where ‘H’ equals either A, T, or C) and predominantly concerns transposons and other repetitive DNA elements, though methylation of expressed genes has also been frequently described. The overall DNA methylation pattern in plants thereby appears to be quite static and maintained in a multigenerational manner with common exceptions such as genome-wide demethylation during development. Stable maintenance of methylation insures, once established, that transposons remain in a silenced state and preserves cell type identity (Law and Jacobsen 2010, Law et al. 2011).

One striking and well known epigenetic example of DNA methylation is the change from bilateral to radial floral symmetry in the toadflax *Linaria vulgaris*. The phenotype of the flower correlates tightly with the degree of DNA methylation at the promoter of the *Lcyc* gene which controls dorsoventral asymmetry, and is inherited over several generations. The *Linaria* variety with radial flowers is extensively methylated at the promoter and thus the *Lcyc* gene is silent, while the variety with demethylated promoter correlates with the active state of the *Lcyc* gene and exhibits the wild-type bilateral flowers. Transitional stages between wild-type and radial flowers are also known and involve different degrees of methylation of the *Lcyc* gene. The radial phenotype is also not always stable and occasionally reverts back to the wild-type phenotype during somatic development (Cubas et al. 1999,

Richards et al. 2012). Another example closely related to the chemotype formation in thyme is the striking phenotypic variation in a mosaic *Eucalyptus melliodora* tree first described in 1990. This tree possesses a branch with a chemically distinct phenotype different to the chemotype exhibited in the rest of the tree, which also increased resistance to insect herbivory. Even though the underlying control mechanism is still unknown it is likely that DNA methylation and histone modifications play an important role in this case. SNPs were found, among others, in two genes which could be involved in the epigenetic regulation of these two 'chemotypes': a 'methyl CpG binding domain 10' like gene and a 'histone deacetylase complex subunit sap18' like gene (Padovan et al. 2013).

Not only somatic mutations but also environmental induction of epigenetic variation can be stably inherited across generations. This could be demonstrated in dandelion (*Taraxacum officinale*) and *A. thaliana* (Richards et al. 2012), and also seems to apply to a certain degree for wild populations of mangrove trees (Lira-Medeiros et al. 2010). Dandelion plants, for example, exposed to different environmental stresses, such as chemical induction of herbivore and pathogen defenses, had considerable methylation variation throughout the genome when compared to control plants, which where, to a high degree, inherited in the next generation (Richards et al. 2012). While exposure of *A. thaliana* plants or *Laguncularia racemosa* mangrove trees to salt stress, resulted in offspring with a greater salt tolerance. In the case of *A. thaliana* it could be demonstrated, by demethylation with 5-azaC, that this enhanced salt tolerance appeared to be controlled by DNA methylation while the mangrove study was performed on field collected material and did not control for environmentally induced epigenetic effects so the exact role of DNA methylation in this case is still unclear (Lira-Medeiros et al. 2010, Richards et al. 2012). The preferences that thyme chemotypes show in connection to soil and temperature might be partially correlated with chemotype specific methylation patterns.

In this study, we obtained no evidence for DNA methylation of the *Tv*TPS1 gene sequence. Three of the restriction enzymes employed in the DNA blotting are sensitive to CpG-methylation (EcoRI, XhoI, and XmaI), while BspHI is sensitive to Dam or Dcm methylation and HindIII is insensitive to any kind of methylation at its restriction site. However, no methylation sensitive or insensitive isoschizomers of the used restriction enzymes were utilized so it is not possible to draw any conclusions about the methylation status of the gene.

5 Plant regeneration from *Thymus vulgaris* leaf explants

Introduction

Genetic transformation is a versatile tool to address questions about the biosynthesis and regulation of essential oils *in planta*. In mint, transformation has been successfully applied to manipulate the production of menthofuran and menthol by overexpression or cosuppression of either precursor-providing genes (Krasnyanski et al. 1999, Mahmoud and Croteau 2001), or enzymes modifying basic monoterpene structures (Mahmoud and Croteau 2003, Mahmoud et al. 2004). In tobacco, transformation was used to investigate the function of a cytochrome P450 candidate involved in cembratriene-diol biosynthesis (Wang et al. 2001) as well as to modify monoterpene biosynthesis by introducing novel genes from other plant species (Lücker et al. 2004a, Lücker et al. 2004b). Using similar approaches, the function of the isolated cytochrome P450 genes from thyme could be investigated and their role in the carvacrol/thymol biosynthesis could be studied by either up- or down-regulating genes responsible for precursor biosynthesis (e.g. γ-terpinene synthase), or by regulating the expression of the cytochrome P450 candidate genes themselves.

Several successful transformation protocols for members of the Lamiaceae family have been established, including mint (Diemer et al. 1998, Niu et al. 1998, Niu et al. 2000), perilla (Kim et al. 2004), and lavandin (Dronne et al. 1999b). However, to the best of our knowledge, no such protocol is available for species of the genus *Thymus*. Before an *Agrobacterium tumefaciens*-mediated transformation protocol can be developed for *Thymus* sp., several conditions have to be met. For one, it is necessary to have an *A. tumefaciens* strain which infects thyme well. Although the susceptibility of thyme to *A. tumefaciens* infections has not been investigated to date, the strain EHA105/GI, seems to be a suitable candidate, as it has been efficiently used for transformation of many plant species including several members of the Lamiaceae (Diemer et al. 1998, Niu et al. 1998, Dronne et al. 1999b, Kim et al. 2004). Another important point that needs to be addressed is the possible cytotoxicity and/or the harmful influence introduced genes can have when under strong or constitutive expression. This is of special importance in thyme, since the isoprenoid pathway supplies precursors not only for the mono- and sesquiterpene biosynthesis but also for other compounds essential for plant growth and fitness (Mahmoud and Croteau 2001, Gog et al.

2005, Aharoni et al. 2006), and diversion of too much isoprenoid flux to other compounds might negatively impact growth and development. To alleviate this problem one could use an inducible promoter, rather than a strong, constitutive one, or one could use trichome-specific promoters to limit gene expression to the glandular trichomes. In this regard, it should be noted that several promising trichome-specific promoters have been characterized in recent years (Gutiérrez-Alcalá et al. 2005, Liu et al. 2006, Shangguan et al. 2008, Ennajdaoui et al. 2010, Wang et al. 2013). Finally, last but not least, genetic transformation requires an efficient regeneration system, which is the focus of this study.

Although no regeneration system, suitable for a transformation protocol, has been established for thyme, several successful micropropagation protocols for members of the genus *Thymus* have been described to date, including *T. caespititius* (Mendes et al. 2013), *T. longicaulis* (Ozudogru et al. 2011), *T. lotocephalus* (Coelho et al. 2012), *T. mastichina* (Mendes and Romano 1999, Fraternale et al. 2003), *T. piperella* (Sáez et al. 1994), *T. sipyleus* (Erdağ and Yürekli 2000), *T. spicata* (Daneshvar-Royandezagh et al. 2009), and *T. vulgaris* (Furmanowa and Olszowska 1980, Lê 1989, Furmanowa and Olszowska 1992, Farag et al. 2001, Ozudogru et al. 2011). However, these protocols were mostly designed to facilitate the multiplication of pharmaceutically desirable individual plants for field cultivation, rather than to lay the foundation for a successful transformation of thyme. Yet, some studies have focused on the callus induction from leaf explants of *Thymus vulgaris*, and/or the influence of growth regulators on the production of secondary metabolites in the established *in vitro* plantlet cultures and callus cultures (Sugisawa et al. 1988, Tamura et al. 1993, Farag et al. 2001, Affonso et al. 2009, Mendes et al. 2013).

The aim of this study was to establish a regeneration protocol for thyme, which could be used as a starting point for the development of an *A. tumefaciens* mediated transformation protocol. The experimental set up closely followed the regeneration protocol established for lavandin (*Lavandula* x *intermedia* Emeric ex Loiseleur), a member of the mint family (Dronne et al. 1999a). To ensure regeneration, the primary experimental setup was comprised of three successive media. The initial medium designed to induce callus formation in leaf explants, was followed by a caulogenesis medium to promote shoot induction, and finally a rooting medium, after which plantlets were acclimatized to greenhouse conditions.

Material and Methods

Plant material and chemicals

T. vulgaris variety 'Deutscher Winter' seeds (Quedlinburger Saatgut GmbH, Quedlinburg, Germany) were purchased from a local garden center and surface sterilized with 70 % [v/v] EtOH for 3 min and 10 min in 2 % [w/v] active chloride solution, then rinsed three times in sterilized water. After the third rinse the water was changed again and the seeds were left submersed over night at 4 °C to allow the mucus layer of the seed coat to fully expand. Approximately 30 seeds were placed in Petri dishes on 0.5 x Murashige and Skoog basal medium (Murashige and Skoog 1962), solidified with 0.6 % [w/v] plant agar (Duchefa Biochemie B.V., Haarlem, Netherlands). The medium was adjusted to pH 5.8 and then sterilized by autoclaving at 121 °C for 20 min. Seeds were germinated at 25 °C under an 11 / 13 h (day/night) photoperiod (60 µmol s^{-1} m^{-2}). Young leafs of 14-day-old seedlings were used as explants for all the regeneration experiments.

The protocol for plantlet regeneration used a series of three to four culture media, starting with callus development, followed by caulogenesis, shoot elongation, and finally rooting. The basal media in all these steps consisted of 0.5 x Murashige and Skoog (1962) salts supplemented with 1 % [w/v] sucrose, solidified with 0.6 % [w/v] plant agar (all from Duchefa Biochemie B.V.), and adjusted to pH 5.8 with NaOH prior to autoclaving (20 min at 120 °C). The employed growth regulators where either added prior to sterilization, in the case of IAA (indole-3-acetic acid solubilised in 1 N NaOH and stored at -20 °C) and BAP (6-benzylaminopurine solubilised in 1 N NaOH and stored at -20 °C), or filter-sterilized and added to low warm media in the case of IBA (indole-3-butyric acid solubilised in EtOH and stored at -20 °C) and GA_3 (gibberellin A_3 solubilised in 70 % EtOH and stored at 4 °C).

Young leafs of 14-day-old seedlings, grown under sterile conditions, were aseptically cut, placed in 100 x 20 mm petri dishes (~20 explants per dish) with the abaxial side in contact with the medium, and sealed with Parafilm®. The cultures were kept at 25 / 23 °C (day/night) under cool-white fluorescent light (18 µmol m^{-2} s^{-1}) with an 11 h photoperiod in Percival CU36L5 growth chambers (Percival scientific Inc., Perry, USA). Three callus induction media as well as six caulogenesis induction media were tested. The callus induction media (C1 – C3) consisted of basal medium supplemented with different combinations of cytokinin BAP and auxin IAA (see Table 5-1). The callus development was scored 7 and 14

days after placement on the medium, by counting the number of explants which had developed callus. The caulogenesis induction media (B2, B4, B6, B7, B8, and B9) consisted of basal medium supplemented with different amounts of cytokinin BAP (see Table 5-2 and Table 5-3). To facilitate shoot elongation basal medium supplemented with 1 µM GA_3 was tested, while root induction was promoted by using basal medium supplemented with 1 µM IBA and sucrose concentrations of 0 – 2 % [w/v].

Results

Callus formation from leaf explants

The influence of growth regulators on callus formation and development from leaf explants was investigated using three different concentrations of the synthetic cytokinin, 6-benzylaminopurine (BAP) in combination with the auxin, indole-3-acetic acid (IAA), named C1 – C3 (Table 5-1).

Aseptic seeds of *Thymus vulgaris* germinated well within 14 days on Murashige & Skoog (MS) basal medium and provided sufficient explant material for the experiments. MS medium supplemented with different combinations of BAP and IAA induced callus formation in all tested media in contrast to leaf explants cultured on MS medium devoid of plant growth regulators, which did not significantly develop callus and did not initiate adventitious buds. The callus developed mainly from the cut edges of the leaf explants, and was of a fresh light green color with a compact caulogenetic appearance (Table 5-1).

Medium	IAA [µM]	BAP [µM]	Number of explants	Explants with callus (7 days)			Explants with callus (14 days)		
				no.	[%]	mean	no.	[%]	mean
C1	4.5	4.5	115	7	6.09	29.57 (10.33)	21	18.26	41.14 (15.23)
			94	5	5.32		33	35.17	
			40	14	35.00		28	70.00	
			332	188	56.63		n.a.	n.a.	
			754	338	44.83		n.a.	n.a.	
C2	4.5	9.0	95	7	7.37	14.06 (8.91)	24	25.26	39.02 (8.21)
			97	3	3.09		37	38.14	
			41	13	31.71		22	53.66	
C3	9.0	4.5	95	13	13.68	19.07 (7.64)	53	55.79	55.88 (4.33)
			64	6	9.38		31	48.43	
			41	14	34.15		26	63.42	

Table 5-1: Effect of different indole-3-acetic acid (IAA) and 6-benzylaminopurine (BAP) concentration combinations on callus formation of *T. vulgaris* on half-strength MS basal medium. For each medium, 3 to 5 trials were conducted and the results were scored after 7 and 14 days of culture. Standard error is shown in parenthesis. Abbreviations are as follows: Number of explants with callus (no.), percent of explants with callus (%), mean percent of callus formation (mean), data not available (n.a.).

Leaf explants which formed callus within 7 days on C1-medium ranged from 5.3 – 56.6 %, with a mean of 29.6 %, and after 14 days mean callus formation was 41.1 % with a

range of 18.3 – 70.0 %. Callus formation on C2- and C3-medium after 7 days ranged from 3.1 – 31.7 % and 9.4 – 34.2 %, with a mean of 14.1 and 19.1 %, respectively. After 14 days the mean callus formation was 39.0 % with a range from 25.3 – 53.7 % on C2-medium, and 55.9 % with a range of 48.4 – 63.4 % on C3-medium (Table 5-1). Healthy green explants, both with and without callus, as well as healthy and green-looking callus excised from declining explants were transferred to new medium to induce shoot formation.

After 14 days on callogenesis medium (medium for callus formation), an onset of explant decline could be observed. There appeared to be two different types of decline. The explants representing the first type never developed callus and were of a brown color with a putrefacient-mushy appearance and texture. The explants of the second form first developed healthy green callus, and later on started to turn brown and displayed a dried-out and stiff texture, which could be easily crumbled to smaller pieces with forceps. The callus on this latter form turned gradually brown only after the entire leaf explant had turned brown.

Shoot initiation and proliferation

Calli obtained from the different callogenesis media were transferred to caulogenesis medium (B2, B4, B6, and B8), consisting of MS basal medium supplemented with different BAP concentrations (4.5, 9.0, 18.0, and 35.5 μM BAP), to determine the influence of callogenesis medium on subsequent bud formation, as well as to determine the best conditions for shoot regeneration (Table 5-1 and Table 5-2).

The highest percentage of shoot production observed in individual experiments, after transfer from callogenesis to caulogenesis medium, was 17.4 % on B4 medium after preculture on C1 medium, 57.9 % on B8 medium after preculture on C2 medium, and 13.3 % on B6 medium after preculture on C3 medium, respectively. Highest mean percentage in shoot production observed was 13.0 % on B8 medium after preculture on C1 medium, 34.0 % on B8 medium after preculture on C2 medium, and 11.2 % on B6 medium after preculture on C3 medium, respectively.

Explants which had been precultured on C1 medium showed the lowest shoot induction percentage on B2 medium with an average of 2.2 %, while explants on B6 and B8 medium both had a shoot production average of 10.1 and 13.0 %, respectively. The highest shoot induction in an individual experiment was observed on B4 medium with 17.4 %, which could not be reproduced in the following experiment. When C2 had been used as callogenesis

C-Medium	B-Medium	BAP [µM]	Number of explants	Explants with shoot		
				no.	[%]	mean
C1	B2	4.5	23	1	4.35	2.17
			11	0	0.00	
	B4	9.0	23	4	17.39	8.70
			11	0	0.00	
	B6	18.0	24	3	12.50	10.10
			13	1	7.69	
	B8	35.5	25	4	16.00	13.00
			10	1	10.00	
C2	B2	4.5	23	0	0.00	0.00
			11	0	0.00	
	B4	9.0	23	1	4.35	7.17
			10	1	10.00	
	B6	18.0	20	4	20.00	10.00
			10	0	0.00	
	B8	35.5	19	11	57.89	33.95
			10	1	10.00	
C3	B2	4.5	15	0	0.00	6.25
			8	1	12.50	
	B4	9.0	15	0	0.00	0.00
			10	0	0.00	
	B6	18.0	15	2	13.33	11.21
			11	1	9.09	
	B8	35.5	15	2	13.33	6.67
			11	0	0.00	

Table 5-2: Effect of different 6-benzylaminopurine (BAP) concentrations and influence of callogenesis medium on shoot formation on leaf explants of *T. vulgaris* on half-strength MS basal medium. Callogenesis was carried out on C1- (4.5 µM IAA and 4.5 µM BAP), C2- (4.5 µM IAA and 9.0 µM BAP), or C3-medium (9.0 µM IAA and 4.5 µM BAP) for one week before explants were transferred to different caulogenesis media (B-medium). For each caulogenesis medium, two trials were conducted and the results were scored after 6 weeks of culture. Abbreviations are as follows: Number of explants with shoots (no.), percent of explants with shoots (%), mean percentage of shoot formation (mean).

medium, no shoot induction could be observed on B2 medium, while an average of 7.2 % was observed on B4 medium. Explants on B6 and B8 medium showed shoot induction of 20.0 and 57.9 % in individual experiments, but failed to achieve similar percentages in the following experiments. Thus the average percentage for shoot induction was 10.0 and 34.0 % for B6 and B8, respectively. When C3 had been utilized as callus induction medium, no shoot

induction could be observed on B4 medium. While B2 and B8 medium both had an average shoot production percentage of 6.2 and 6.7 %, respectively. On B6 medium the highest shoot induction could be observed with an average of 11.2 %.

Higher concentrations of BAP in the caulogenesis medium appeared to be beneficial for shoot induction in thyme. To fine-tune the optimal BAP concentration for the caulogenesis medium, two additional media, named B7 (27.0 µM BAP) and B9 (31.5 µM BAP), respectively, as well as the previously mentioned B6 and B8 media were compared to each other (Table 5-3). Explants that had been cultured on C1 medium to induce callogenesis were transferred to B6 – B9 media, to induce shoot formation. The overall shoot induction percentage was lower than in the earlier caulogenesis experiments with 4.9 % on B6 medium, 8.9 % on B7 medium, 9.1 % on B8 medium, and 4.4 % on B9 medium, respectively.

Medium	BAP [µM]	Number of explants	Explants with callus (2 weeks)			Explants with shoots (6 weeks)		
			no.	[%]	mean	no.	[%]	mean
B6	18.0	86 196	35 70	40.70 35.71	38.21	4 10	4.65 5.10	4.88
B7	27.0	110 185	42 81	38.18 43.78	40.98	7 21	6.36 11.35	8.86
B8	35.5	102 195	56 61	54.90 31.28	43.09	8 20	7.84 10.26	9.05
B9	31.5	115	29	25.22	25.22	5	4.35	4.35

Table 5-3: Effect of different 6-benzylaminopurine (BAP) concentrations on callus and shoot formation on leaf explants of *T. vulgaris* on half-strength MS basal medium. Callogenesis was carried out on C1-medium (4.5 µM indole-3-acetic acid (IAA) and 4.5 µM BAP) for one week before explants were transferred to different caulogenesis media. For each caulogenesis medium two trials were conducted, with the exception of B9, and the results were scored after 2 and 6 weeks of culture. Abbreviations are as follows: Number of explants with callus or shoots (no.), percent of explants with callus or shoots (%), mean percentage of callus or shoot formation (mean).

Shoot elongation with the help of 1 µM GA_3 in basal media, as had been successfully used for lavandin (Dronne et al. 1999a), was not further investigated, as explant decline was severe. After the shoots were well developed on their respective caulogenesis medium, they were transferred onto a rooting medium to induce adventitious root formation.

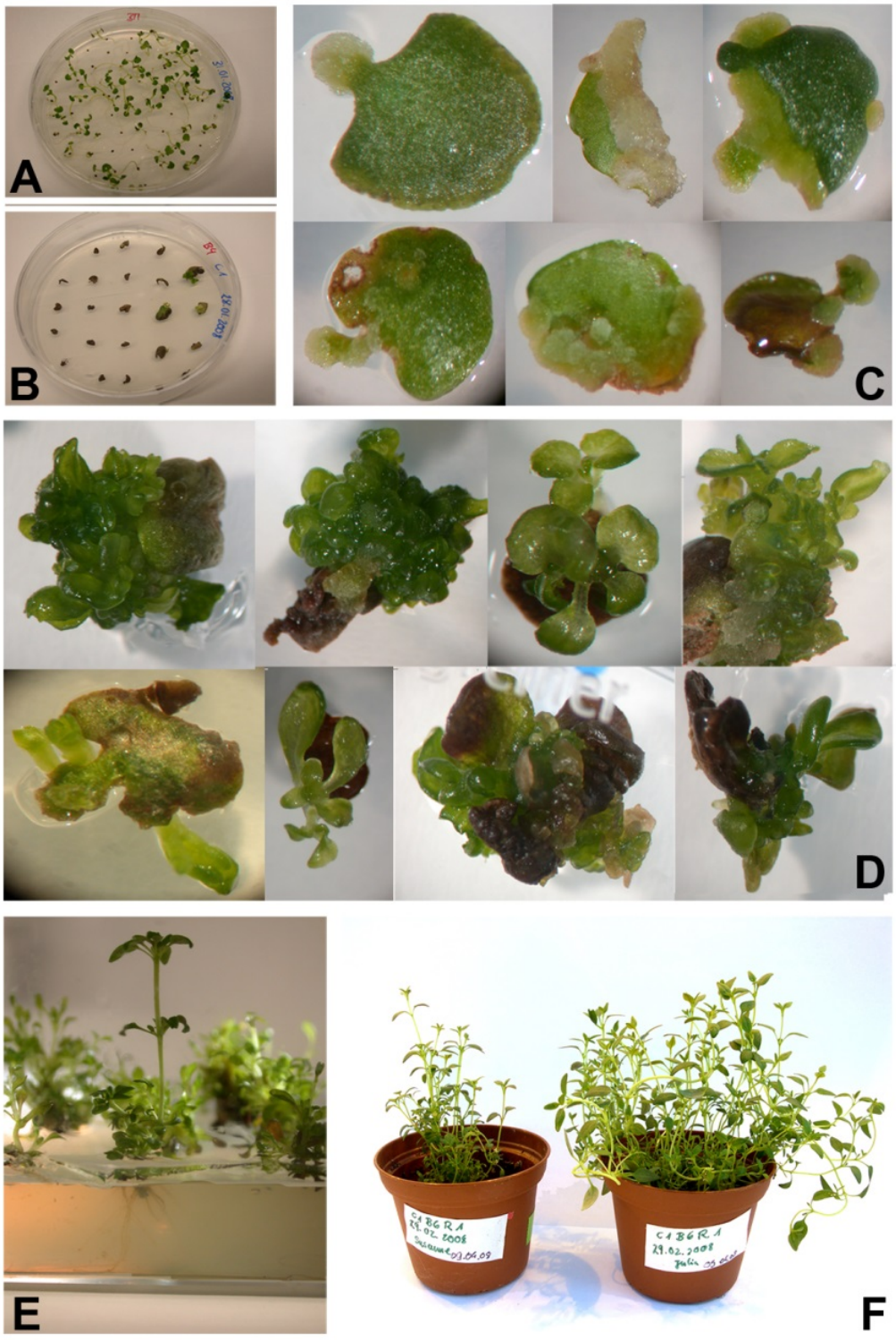


Figure 5-1: Plant regeneration from leaves of *T. vulgaris* var. 'Deutscher Winter'. (A) Seedlings on basal medium (B) leaf explants on caulogenesis medium (C) callus formation on leaf explants (D)

embryo and shoot formation on leaf explants (E) rooted shoots (F) plantlets after acclimatization in the greenhouse.

Adventitious root formation and plantlet acclimatization

Even though rooting is often a difficult phase in the process of micropropagation, root formation of *Thymus vulgaris* explants was successfully induced on half-strength MS medium supplemented with 1 µM IBA. Independent of the tested sucrose concentrations in the medium (0 %, 0.5 %, and 1 % sucrose), the rooting success rate was 100 % after 4 – 6 weeks of culture without subculture on fresh medium. After transfer to soil, greenhouse acclimatization was gradually carried out over the course of two weeks, with a survival rate of 100 % (Figure 5-1). The plantlets grew with a normal morphology, and the appearance and size of the leaves were uniform, and similar to those of seedlings grown under normal conditions in the greenhouse.

Discussion

Traditional micropropagation systems, which exploit the morphogenic potential of existing meristems within the plant, are only of limited use in plant transformation. Therefore explants which undergo direct organogenesis or somatic embryogenesis are typically employed. In this indirect method of regeneration via somatic embryogenesis, somatic cells pass through a number of characteristic evens – dedifferentiation of cells, activation of cell division, as well as reprogramming of their physiology, metabolism, and gene expression patterns – resulting in the formation of somatic embryos capable of regenerating plants. Similar to traditional micropropagation, have environmental factors, such as carbon source in the medium, temperature, light quality and photoperiod, as well as plant growth regulators a great influence on the adaption, survival, and developmental transition of plant tissue. In addition to this is the potential of somatic embryogenesis known to vary with plant species, among genotypes within a species, and with the physiological state of the tissue background (Deo et al. 2010).

Exogenous auxin and cytokinin supplementation via the culture medium is often used to initiate and regulate somatic embryogenesis, as these plant growth regulators affect the endogenous hormone concentrations. Depending on plant species and explant type, supplementation of exogenous auxin or cytokinin can be redundant, as endogenous auxin or cytokinin levels are sufficient. However, callus formation is commonly induced by supplying low concentration of exogenous cytokinin and auxin. Subsequent shoot induction, meanwhile, is often promoted by a high cytokinin concentration while auxin is often omitted from the medium, as the presence of auxin promotes callus proliferation and inhibits cell differentiation. The removal of exogenous auxin also facilitates the establishment of an endogenous auxin gradient which is thought to be essential for the establishment of bilateral symmetry during plant embryogenesis (Jiménez 2005, Deo et al. 2010).

Callus formation on explants

Previous attempts have been made to micropropagate species of *Thymus* and other genera of the Lamiaceae, and several studies have investigated the effects of growth regulators on the development of callus from *T. vulgaris* explants and the essential oil production therein (Sugisawa et al. 1988, Tamura et al. 1993, Farag et al. 2001). Similar to these earlier studies was it possible for us to induce callus formation on leaf explants,

obtained from *T. vulgaris* seedlings grown under sterile condition, by using callogenesis medium supplemented with the natural occurring auxin indole-3-acetic acid (IAA) and the synthetic cytokinin analogue 6-benzylaminopurine (BAP) in varying concentrations and combinations (Table 5-1 and Figure 5-1). The initiation and development of callus on leaf explants on medium devoid of growth regulators was, as expected, not significant and no adventitious buds were initiated. This limited callus formation demonstrates, however, that endogenous auxin and cytokinin levels of these explants are relatively high. The addition of exogenous auxin and cytokinin to the medium strongly stimulated callogenesis in all tested media, also as expected, though the induction was not as successful as callus induction described for lavandin (*Lavandula* x *intermedia*) (Dronne et al. 1999a), *Salvia officinalis*, and *S. fruticosa* (Kintzios et al. 1999). In *S. officinalis*, callus induction could also be successfully initiated by applying only the cytokinin-like compound thidiazuron (TDZ) with no additional auxin supplementation to the medium (Tawfik and Mohamed 2007). Compact lime green caulogenetic calluses formed mostly at the base of the cut surfaces of the *T. vulgaris* leaf explants and in some cases on the whole leaf. From the three callogenesis media tested in this thesis, produced C3-medium (9.0 µM IAA and 4.5 µM BAP) callus at a rate of 56 %, while both C1-medium (4.5 µM IAA and 4.5 µM BAP) and C2-medium (4.5 µM IAA and 9.0 µM BAP) were slightly less efficient with 41 % and 39 % of callus formation on leaf explants, respectively. However, it should be noted that the highest callus induction observed in a single experiment was 70 % on C1-medium, indicating that callus induction will benefit significantly from additional callogenesis medium optimization. In contrast to our experimental setup reported Sugisawa and coworkers satisfactory callus induction on *T. vulgaris* leaf explants using MS medium supplemented with 28.6 µM NAA (1-naphthaleneacetic acid) and 46.5 µM KIN (kinetin or N^6-furfuryladenine) while the best subculture conditions for callus were found to be either 5.7 µM NAA plus 4.7 µM KIN or 4.5 µM 2,4-D (2,4-dichlorophenoxyacetic acid) plus 4.7 µM KIN (Sugisawa et al. 1988, Tamura et al. 1993). In addition to this was a combination of 4.5 µM 2,4-D and 4.7 µM KIN successfully used to induce callus formation by Farag et al. (2001), while the highest subsequent leaf callus growth was reported to be on medium supplemented with 4.5 µM 2,4-D and 9.4 µM KIN, and the lowest growth rate on medium supplemented with 5.4 µM NAA and 8.8 µM BAP. These results, together with our findings, demonstrate that callus induction in *T. vulgaris* explants is readily achieved by utilizing a variety of different auxin and cytokinin types in different proportions and concentrations, making this plant a promising candidate for the establishment of a practical transformation protocol. In this regard it should

be noted that the synthetic auxin analogue 2,4-D is highly effective and one of the most frequently used auxins to induce callogenesis, followed in popularity by NAA, while IAA or IBA (indole-3-butyric acid) are used less frequently (Jiménez 2005). Several different Lamiaceae protocols, next to the above mentioned thyme protocols, have applied 2,4-D successfully to induce callus formation on explants and/or promote subsequent callus growth. So was in lavandin 99.3 % callus formation observed on leaf explants cultivated on medium supplemented with 4.5 µM 2,4-D and 4.5 µM BAP, while supplementation with the auxin NAA lead to a maximum of 98.0 % callus formation (9.0 µM NAA and 4.5 µM BAP) (Dronne et al. 1999a). The high efficiency of 2,4-D is attributed to its noted ability to significantly increase the endogenous auxin levels, which can become in later stages of embryogenesis counterproductive. To induce caulogenesis, as mentioned before, exogenous auxin has often to be removed from, or greatly reduced in the culture medium, so as not to hinder the development of a polar auxin gradient. However, even with 2,4-D removed from the medium, endogenous auxin levels often continue to remain undesirably high, consequently inhibiting embryogenesis (Jiménez 2005, Deo et al. 2010). This was also demonstrated in lavandin, where bud formation was only 5 % for explants previously grown on 2,4-D containing medium, in contrast to 80 % for explants previously grown on NAA containing medium (Dronne et al. 1999a). The same phenomenon was observed in *S. fruticosa* and *S. officinalis*, where growth regulators had to be omitted completely from medium to continue the maturation process of adventitious shoot formation (Kintzios et al. 1999).

The type, concentration and combination of plant growth regulators are all important factors influencing callus formation and proliferation in plant tissue cultures. Farag et al. (2001), for instance, found, depending on the growth regulators used, different auxin:cytokinin ratios beneficial for the callus proliferation. The group of Sugisawa, which reported satisfactory proliferation of thyme callus on medium supplemented with 28.6 µM NAA and 46.5 µM KIN, reduced the supplemented growth regulators greatly for callus subculture to 5.7 µM NAA (or 4.5 µM 2,4-D) plus 4.7 µM KIN, thereby effectively altering the auxin:cytokinin ratio from 1:2 to 1:1 (Sugisawa et al. 1988, Tamura et al. 1993). Our results indicate that a 1:1 or 2:1 ratio of IAA to BAP results in a slightly higher callus formation than a 1:2 ratio, which agrees with the observations of Farag et al., as well as the results of Tamura et al. which tested, among others, different NAA:BAP ratios. Both, 2,4-D and NAA are synthetic auxin analogues, however NAA is structurally related to the natural

occurring IAA and thought to have a more IAA-like activity than 2,4-D, which makes the comparison of above mentioned results feasible. The available information on callus proliferation in *T. vulgaris* in specific, and Lamiaceae in general, indicates that this plant family readily forms callus from different explant types and that a wide variety of growth regulators in different concentrations and combinations are effective. The results from initial callus induction on explants from lavandin (Dronne et al. 1999a), *S. fruticosa*, and *S. officinalis* (Kintzios et al. 1999, Tawfik and Mohamed 2007) also demonstrate that the growth regulator concentration together with the auxin:cytokinin ratio are very important control factors for optimizing a protocol for somatic embryogenesis in a member of the Lamiaceae plant family.

It is interesting to note that Kintzios et al. (1999) showed that light conditions as well as growth regulator concentration both significantly affect callus induction and explant decline in *S. officinalis* and *S. fruticosa*. In *S. fruticosa* the callus induction was 100 %, and independent of NAA:BAP ratio, growth regulator concentration (10.5 µM and 21.0 µM), and light intensity used, while callus induction was sensitive to light intensity, auxin:cytokinin ratio, and concentration (1.8 µM, 9.0 µM, and 18.0 µM), when 2,4-D and KIN were applied, and ranged from 73 – 100 %. A significant decline of up to 44 % of the *S. fruticosa* explants was observed at a concentration of 10.5 µM NAA and 10.5 µM BAP, both under high (250 µmol m^{-2} s^{-1}) and low (50 µmol m^{-2} s^{-1}) light intensity, while other concentrations and combinations did not lead to significant losses. Similar to this, lead the application of 2,4-D and KIN to the callogenesis medium to an overall higher explant decline, and in particular under low light conditions where decline often reached 100 %. In *S. officinalis*, explants showed light sensitivity in all tested growth regulator combinations and concentrations. The best callus induction parameters for *S. officinalis* were either 1.8 µM 2,4-D and 1.8 µM KIN under low light (100 % callus formation and 0 % explant decline) or 10.5 µM NAA and 10.5 µM BAP under high light (67 and 27 %, respectively) and for *S. fruticosa* either 1.8 µM 2,4-D and 1.8 µM KIN under high light (100 and 25 %, respectively) or 10.5 µM NAA and 21.0 µM BAP under high light (100 and 0 %, respectively). Another callus induction protocol for *S. officinalis*, which utilizes only thidiazuron (TDZ) in the culture media, calls for a one-week incubation phase in total darkness after explant preparation, followed by four weeks under low light regime (40 µmol m^{-2} s^{-1}) (Tawfik and Mohamed 2007). These observations demonstrate adequately that other environmental factors, apart from plant growth regulators can have a significant impact on plant tissue cultures, and also show that these factors

influence one another and subsequently the plant tissue (Deo et al. 2010). It is thus feasible to assume that the low light conditions (18 µmol m^{-2} s^{-1}), which we used in our callogenesis experiments had, in combination with the plant species and the cultivar (*T. vulgaris* var. 'Deutscher Winter'), as well as the growth regulator types and their respective concentrations, a negative impact on either the callus formation and/or the high explant decline we observed. The previously described studies on *T. vulgaris* callus, for example, used light conditions of approx. 27 µmol m^{-2} s^{-1} (2000 Lux) without reported difficulties (Sugisawa et al. 1988, Tamura et al. 1993, Farag et al. 2001), while callus induction in lavandin was successfully accomplished under very low light (6 µmol m^{-2} s^{-1}) (Dronne et al. 1999a).

Shoot initiation and proliferation

Adventitious shoot organogenesis, which is also referred to as caulogenesis, is the process where totipotent plant cells produce a unipolar structure, namely a shoot, whose vascular system is connected to the parent tissue. In contrast, somatic embryogenesis leads to the production of a bipolar structure containing a root/shoot axis, and possesses a vascular system independent from the parent tissue. Both processes depend on several factors, such as endogenous and exogenous growth regulators, particularly auxins and cytokinins, which can be used to direct the differentiation, competence and morphogenesis of the somatic cells (Thomas et al. 2004, Jiménez 2005, Deo et al. 2010, Magyar-Tábori et al. 2010). Usually, not only the quantity of auxins and cytokinins per se, but also the proportions of one to the other, are determining factors for control of cell cycle, cell division, and differentiation (George et al. 2008). In several Lamiaceae, bud regeneration from leaf explants could be successfully achieved by using a single medium, supplemented with an auxin/cytokinin combination, to promote both callogenesis and caulogenesis. In most cases supplementing the medium with a combination of an auxin and a cytokinin lead to the best results (Erdağ and Yürekli 2000, Kim et al. 2004, Skała and Wysokińska 2004, Bouhouche and Ksiksi 2007, Xing et al. 2007), while it was sufficient in *Mentha* x *piperita* to supply only exogenous cytokinins (Niu et al. 1998). Other protocols, such as the one we used, induce callogenesis and caulogenesis separately by using several successive media (Dronne et al. 1999a, Raha and Roy 2003, Tawfik and Mohamed 2007).

Explants transferred from the callus induction media (C1 – C3) to bud induction media supplemented with 4.5 – 35.5 µM BAP but devoid of auxin, formed shoots on all

tested BAP concentrations, although in very low numbers (Table 5-2 and Table 5-3). In addition, the callogenesis medium seemed to have a slight influence on the formation of shoots, as explants from C1 and C2 medium fared better than those obtained from C3 medium. To our knowledge, this is the first report of adventitious shoot formation from leaf explants of *Thymus vulgaris* via indirect morphogenesis.

Bud regeneration was more frequent on caulogenesis media with BAP concentrations ranging from 18.0 – 35.5 µM BAP, than on media with lower concentrations, and a slight difference could be observed depending on the callogenesis medium used (Table 5-3). However, the bud regeneration rate overall was very low and ranged mostly around 10 %. Dronne et al. (1999a) found that bud regeneration for lavandin had an optimal BAP range from 9.0 – 35.0 µM, which also differed slightly depending on the medium used for callus induction, but induced bud formation rates from 60 % to up to 98 %.

High BAP concentrations of 22.2 µM were also reported to induce satisfactory shoot formation rates of 52.3 % in peppermint, but led at the same time to 3.5 % root formation, which was not desirable. A concentration of 8.4 µM TDZ, in contrast, led to 75.4 % shoot induction with no rooting, and was therefore used to establish a successful transformation protocol for peppermint (Niu et al. 1998). In just the same manner as high BAP concentrations have successfully induced bud regeneration in peppermint, lavandin, and thyme, low to medium amounts of BAP (4.5 – 13.5 µM) in combination with low auxin concentration successfully induced bud regeneration in other Lamiaceae, such as *Thymus sipyleus* (Erdağ and Yürekli 2000), *Perilla frutescens* (Kim et al. 2004) and *Lavandula angustifolia* (Xing et al. 2007). The protocol for *Salvia nemorosa*, however, used a cytokinin/auxin combination of 0.9 µM BAP and 2.9 µM IAA (Skała and Wysokińska 2004), while several different media with either NAA plus BAP, or devoid of any growth regulators were used for *S. officinalis* and *S. fruticosa* (Kintzios et al. 1999, Tawfik and Mohamed 2007).

In our study, medium B8 (35.5 µM) induced more bud formation than the other caulogenesis media, but mainly induced the formation of hyperhydric dark green buds (Figure 5-1), from which only some matured into shoots. Several factors are known to correlate with development of hyperhydricity in plant cell cultures, such as the type and concentration of growth regulators, the concentration of ingredients of the culture medium, type and concentration of the gelling agent, low light intensity, the genotype used, and high

humidity (Debergh et al. 1992, George et al. 2008, Picoli et al. 2008). In this study only the amounts of BAP differed between the caulogenesis media used, and hyperhydricity was only observed in explants on medium with the highest BAP concentration. Coelho et al. (2012) also observed an increase in hyperhydric shoots of *T. lotocephalus*, when micropropagated shoots were grown on MS medium supplemented with different BAP concentrations (0.9 - 4.4 μM). This phenomenon was aggravated when auxin (1.1 μM IAA) was added to the medium. While tissue cultures of dicotyledonous plants generally require exogenous cytokinins in the culture medium to sustain growth, as they seem to be unable to synthesize sufficient amounts under *in vitro* conditions, higher concentrations can lead to strange deformations, inhibition of shoot elongation, and the aforementioned hyperhydricity (George et al. 2008, Picoli et al. 2008). Reduced BAP concentrations, as in medium B7 (27.0 μM BAP), helped to avoid hyperhydricity in shoots, but induced simultaneously fewer shoots in explants.

Adventitious root formation and greenhouse acclimatization

Induction of adventitious root formation under *in vitro* conditions is often a critical phase in the plant regeneration process. It is known that the rooting of explants *in vitro* improves the survival and growth rate during the acclimatization process. Since malfunctioning stomata sometimes cause water loss in *in vitro* grown explants at this stage. It is assumed that this water loss can be better compensated by the explants when roots were formed *in vitro* prior to acclimatization (De Klerk 2002). Adventitious root formation itself consists of three successive but interdependent phases: induction, initiation and expression. The induction phase comprises of molecular and biochemical events without visible changes and requires high auxin and/or cytokinin concentrations. The initiation phase is characterized by cell divisions and root meristem formation. The expression phase is defined by intra-stem growth of root primordia and root emergence; auxin can have an inhibitory effect on the root formation at this stage (De Klerk 2002, Li et al. 2009). The environmental and endogenous factors (e.g. phytohormones, mineral nutrients, carbohydrates, and light) influencing root formation are diverse, and vary from species to species and sometimes from genotype to genotype (Li et al. 2009).

Previous studies on *T. vulgaris* micropropagation have shown that rooting of shoots can be readily achieved in this plant species (Lê 1989, Furmanowa and Olszowska 1992, Shetty et al. 1996, Ozudogru et al. 2011). While Furmanowa et al. observed best rooting

results on medium supplemented with 2.5 µM IBA in conjunction with or without 0.2 µM KIN or 0.05 µM 2-isopentyl adenine (2-iP), Lê as well as Shetty et al. achieved high rooting success rates on medium devoid of growth regulators. Ozudogru et al. reported ease of rooting on medium supplemented with either 0 – 5.7 µM IAA, 0 – 4.9 µM IBA, 0 – 5.7 µM NAA, or 0 – 6.1 µM 2,4-D, with the best root apparatus occurring on medium supplemented with 0.3 µM 2,4-D (92.5 % rooting rate, with 19 adventitious roots per shoot). Like Furmanowa et al. and Ozudogru et al., we also did not encounter problems rooting the obtained shoots on medium supplemented with an auxin (1 µM IBA). As mentioned before, high endogenous levels of either auxins and/or cytokinins are necessary to induce adventitious root formation. IAA and IBA are therefore routinely used to promote root formation in plant tissue cultures, by supplying additional exogenous auxin. IBA is known to have a higher root-inducing capacity than IAA, but the response to the type of auxin used is also species dependent (De Klerk 2002, Li et al. 2009). The results indicate that root formation in *T. vulgaris* is readily achieved and that exogenous auxin supplementation is not necessary, as endogenous auxin levels are high enough, but that supplying exogenous auxin can aid root apparatus formation (Lê 1989, Furmanowa and Olszowska 1992, Shetty et al. 1996, Ozudogru et al. 2011).

Another factor which can significantly influence the root induction and formation in tissue cultures is the basic nutrient composition of the medium. The nitrogen source and salt concentrations are of particular interest here, as NH_4^+ ions can have a toxic effect on *in vitro* tissue cultures of higher plants whether it is the only N source or when NO_3^- is present in the medium. Often both NH_4^+ and NO_3^- are employed because together these lead to enhanced nitrogen transport to the shoot. However, the reduction or removal of NH_4^+ from the medium can induce or enhance adventitious shoot and root formation (Lê 1989, Parra and Amo-Marco 1996, Britto and Kronzucker 2002, Bennett et al. 2003). In this regard, it is interesting that several rooting protocols described for Lamiaceae species either use half strength MS medium (Shetty et al. 1996, Dronne et al. 1999a, Mendes and Romano 1999, Kim et al. 2004, Coelho et al. 2012), quarter strength MS medium (Coelho et al. 2012), or modified MS-medium (CMS-medium) (Lê 1989, Sáez et al. 1994, Mendes and Romano 1999) which have NH_4^+ concentrations of 5.15 mM, 2.58 mM, and 2 mM, respectively. However, this study used half-strength MS medium, while others efficiently used either Nitsch & Nitsch medium (Furmanowa and Olszowska 1980, Furmanowa and Olszowska 1992) or full strength MS-medium (Lê 1989, Niu et al. 1998, Daneshvar-Royandezagh et al. 2009, Ozudogru et al.

2011) with NH_4^+ concentrations of 9 mM and 10.3 mM, respectively. The comparison of root formation of thyme explants on CMS-medium (2 mM NH_4^+) and full strength MS-medium (10.3 mM NH_4^+) showed, that although rooting success rate was 100 % for both media, more roots per shoot were formed on CMS-medium (Lê 1989). In general it can be concluded that thyme is quite tolerant to NH_4^+ concentrations in the medium and readily roots on medium with high NH_4^+ concentrations as well as low concentrations, both with or without additional growth regulators. However, continuing to use a lower NH_4^+ concentration in future experiments with thyme, could not only have a positive effect on root length and number of roots per explants (Lê 1989), but also on the overall viability of explants, callus, and shoots (Britto and Kronzucker 2002).

The carbohydrates added to the medium are often considered only as meeting an energetic requirement, rather than a factor influencing adventitious root formation. However, several studies demonstrated the important role carbohydrates can play in the regulation of root formation (Corrêa et al. 2005, Li et al. 2009). We therefore tested three different sucrose concentrations (0 – 1 % sucrose) in the rooting medium, but could not observe any positive or negative influence on root formation in *T. vulgaris*. Furthermore, to the best of our knowledge, no study concerning Lamiaceae regeneration specifically mentions the influence of carbohydrates on root formation in this plant family. In conclusion it can be said that we did not encounter problems in rooting the shoots using half-strength MS-medium supplemented with 1 µM IBA and 0 – 1 % sucrose. This coincides with earlier studies on *T. vulgaris*, and verifies that this plant species is easy to root under different conditions. In addition, greenhouse acclimatization of rooted explants was readily achievable, with a survival rate of up to 100 %.

Lethal browning

Shoot tips and single-node stem segments from *T. vulgaris* plants previously grown in the greenhouse gradually turned brown on the medium within 1 – 2 weeks after surface sterilization and died. The same problem occurred, although delayed and less severe, when leaf explants from sterile grown seedlings were placed on the medium. This phenomenon caused a high loss of material during the regeneration experiments and could not be significantly reduced by frequent subculture to new medium or increase of sucrose and salt concentrations in the media. Numerous studies have reported tissue browning and necrosis leading to poor plant regeneration *in vitro* (Dan 2008). For example, this phenomenon is a

well described problem in banana micropropagation, and has also been observed in several Lamiaceae, including *Rosmarinus officinalis* (Misra and Chaturvedi 1984), *Hyssopus officinalis* (Lê 1987), *Salvia officinalis* and *S. fruticosa* (Kintzios et al. 1999), *Ocimum sanctum* (Shahzad and Siddiqui 2000), *Scutellaria baicalensis* (Fang et al. 2008), and *Thymus sipyleus* (Erdağ and Yürekli 2000). In *T. vulgaris* browning is reported to be the main reason for the loss of ca. 30 % of the initial plant material for micropropagation and only 2 – 3 % loss was attributed to contamination (Lê 1989).

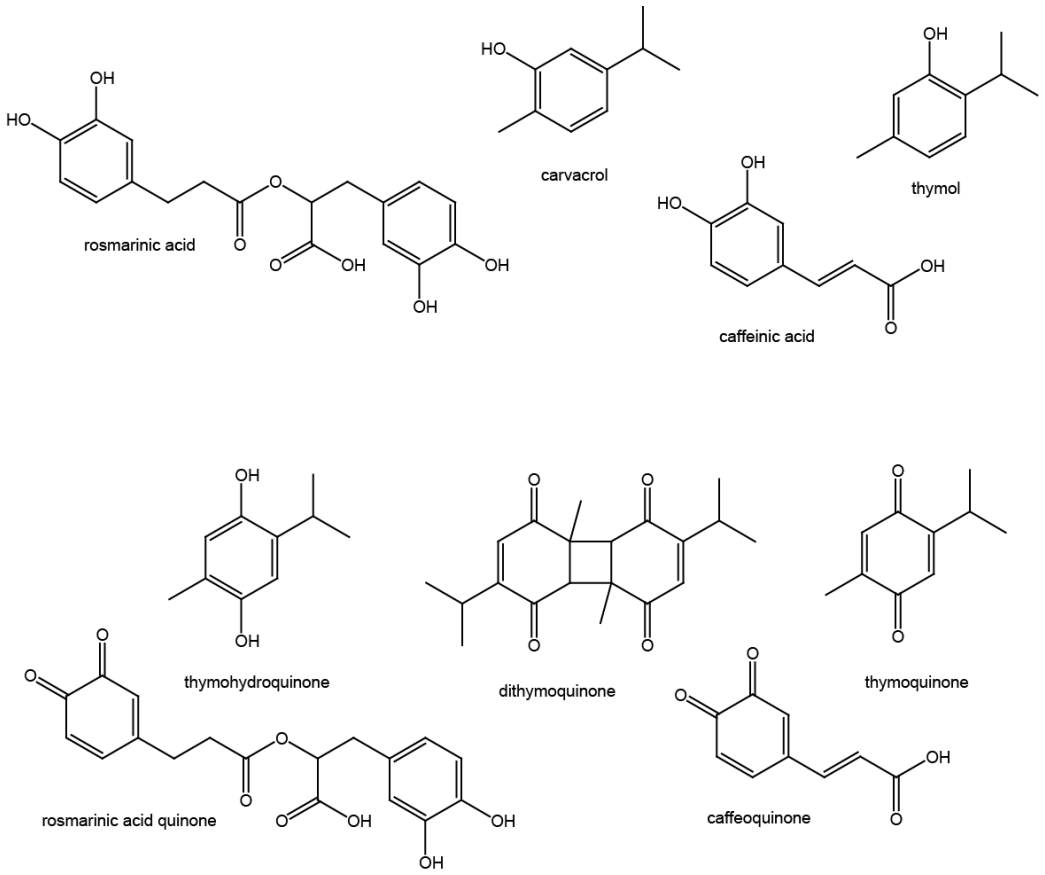


Figure 5-2: Phenolic compounds commonly found in *T. vulgaris* and their corresponding quinones.

Various factors have been suggested as the cause of lethal browning and shoot tip necrosis, including the exudation of phenolics and their subsequent oxidation, prolonged subculture, drop in pH of the medium, and deficiency of calcium (Martin et al. 2007). In particular the deleterious effects of oxidized phenols can be a serious problem in plant micropropagation and is a known obstacle associated with the *in vitro* establishment of monocotyledonous and woody plant species (Kim et al. 2007). It is likely, that the browning

we observed in *T. vulgaris* tissue culture was induced by oxidization of phenols (Figure 5-2). Phenols can be oxidized by polyphenol oxidase (PPO), a copper-containing oxidase, which catalyzes the O_2-dependent oxidation of catechols to their corresponding quinones (Figure 5-3). These quinones are highly toxic compounds to plant tissues (Dicko et al. 2002, Bittner 2006, Fang et al. 2008). Apart from PPO, other oxidative enzymes such as peroxidases (POD), which are heme-containing enzymes that catalyze the conversion of H_2O_2 to water using phenolic compounds as hydrogen donors, and phenylalanine ammonia lyase (PAL) also contribute to wound induced browning. Tabiyeh et al. (2006) report that PAL, which is a branch point enzyme in the phenylpropanoid pathway, has a profound impact on the progress of browning and necrosis of plant tissues. By generating H_2O_2 for the oxidation of phenolic compounds, PPO may act synergistically with POD in enzymatic browning. Both enzymes play an important role in plant defense against pathogens and pests by oxidizing endogenous phenolic compounds into toxic quinones. The resulting quinones may then undergo non-enzymatic auto-polymerization or covalent bonding to proteins and carbohydrates to produce colored compounds. These compounds may constitute a physical as well as chemical barrier against biotic and abiotic stresses (Dicko et al. 2002, Bittner 2006).

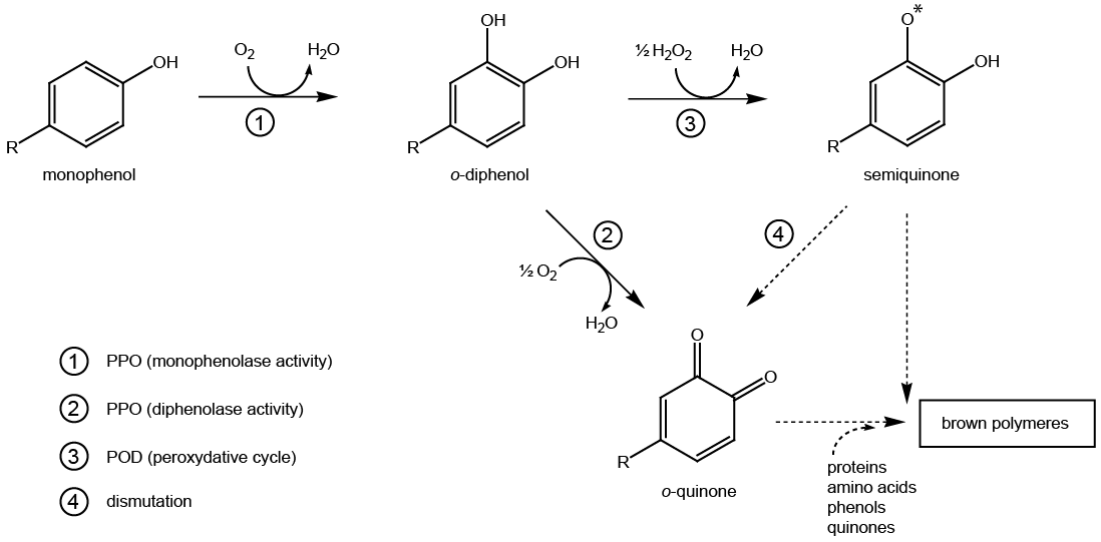


Figure 5-3: Reactions involved in the formation of brown polymers. Polyphenol oxidase (PPO) hydroxylases monophenols to diphenol, and oxidases diphenols to quinones. Peroxidases (POD) use H_2O_2 to produce semiquinones.

Several methods have been described in the literature to alleviate browning of explants and callus cultures caused by oxidation of phenolics and formation of quinones. The

focus of these methods is either reducing the amount of phenolics or inhibiting their oxidation. (i) Phenol-like substances that are exuded by cultured plants can be adsorbed by adding polyvinylpyrrolidone (PVP) or activated charcoal to the culture medium, or using the latter in solutions when preparing explants (Misra and Chaturvedi 1984, Kim et al. 2007, Fang et al. 2008, Krishna et al. 2008, Ko et al. 2009). (ii) Prior to exudation, the production of polyphenols can be influenced by regulating the nitrogen source in the culture medium. It could be shown that the omission of NH_4NO_3 from the medium enhanced the growth and polyphenol production of *Quercus acutissima* (Tanaka et al. 1995) and the polyphenol production of *Sapium sebiferum* (Neera et al. 1992) callus cultures, while the addition of NH_4NO_3 reduced the level of polyphenolics and as a consequence reduced the browning of explants of *Rosmarinus officinalis* (Misra and Chaturvedi 1984). However, both Lê and Sáez (Lê 1989, Sáez et al. 1994) reported better micropropagation on modified MS media with reduced NH_4^+, where $NH_4H_2PO_4$ and $Ca(NO_3)_2$ were used instead of NH_4NO_3, KH_2PO_4 and $CaCl_2$ for *Thymus vulgaris* and *T. piperella*, respectively. (iii) Light was shown to have an aggravating effect on browning by comparison of the level of *in vivo* phenolics, phenol exudation *in vitro,* and activities of oxidative enzymes in etiolated and non-etiolated control plants (Krishna et al. 2008). An effect of light intensity in combination with different growth regulators on lethal browning could also be observed in *Salvia officinalis* and *S. fruticosa* leaf callus cultures (Kintzios et al. 1999). (iv) The activity of some PPO themselves can be inhibited by certain phenolics, such as phloroglucinol and resorcinol which not only prevents quinone production, but also protect auxin from peroxidase-catalyzed oxidation in culture medium (Kim et al. 2007). (v) Finally, the addition of filter-sterilized antioxidants, such as ascorbic acid, salicylic acid, phytic acid, citric acid, or cysteine to the culture media after autoclaving or using them in solutions when preparing explants, was reported to have a positive effect on alleviating lethal browning (Misra and Chaturvedi 1984, Lê 1989, Kintzios et al. 1999, Dan 2008, Fang et al. 2008, Krishna et al. 2008, Ko et al. 2009). Ascorbic acid is a well-known inhibitor of PPO activity, though the mode of action in culture media is not yet understood, since it was shown that ascorbic acid and its oxidation product, dehydroascorbic acid, rapidly decay and disappear from the medium within 12 h after addition (Elmore et al. 1990). Even though the mode of action is not understood, the positive effect that ascorbic acid and other antioxidants have in plant culture media has been repeatedly observed, and it has been reported that antioxidants not only prevent or alleviate lethal browning but also help to increase the number of plantlets and promote callus growth (Tawfik and Mohamed 2007, Dan 2008, Fang et al. 2008, Ko et al. 2009). Antioxidants have also been successfully used in

A. tumefaciens-mediated transformation processes to reduce browning and necrosis in transformed tissue by eliminating pathogen induced reactive oxygen species, and also to promote tissue regeneration and the frequency of transformation (Dan 2008).

6 General Discussion

The main aim of the research presented in this thesis was to gain deeper insight into the regulatory mechanisms underlining the monoterpene polymorphism in *Thymus vulgaris*. For this purpose the terpenoid blend of plants obtained from wild populations in southern France was analyzed and their chemotype affiliation verified (Chapter I). The isolation and characterization of genes involved in the production of major essential oil components was then carried out (Chapter II) to facilitate the investigation of the molecular-genetic basis of the *T. vulgaris* polymorphism (Chapter III). Since the ability to genetically transform thyme would be a desirable tool to prove some of the molecular control mechanisms identified in this thesis, a regeneration protocol for *T. vulgaris* was established as a prerequisite for transformation (Chapter IV).

Thyme chemotypes have more than one level of variation

GC-MS analysis of organic extracts of young expanding leaves of various thyme collections identified them as representatives of the known *T. vulgaris* chemotypes (Chapter I). Since previous observations and controlled crosses had shown that the chemotypes are genetically determined and are not influenced by environmental stress, they were analyzed after being grown under greenhouse conditions. The investigated plants could be identified as distinct chemotypes with varying amounts of the characteristic monoterpenoids of other chemotypes. For example, the non-phenolic chemotypes G_1 (geraniol), A_{10} (*α*-terpineol), and U_6 (thujanol), have essential oil compositions that contain only trace levels of the denominating monoterpenoids of other chemotypes. However, the non-phenolic chemotype L_{48} (linalool) as well as the phenolic chemotypes C_{10} (carvacrol), T_2 (carvacrol/thymol), and T_{28} (carvacrol/thymol) have noteworthy amounts of monoterpenoids denominating other chemotypes. The linalool chemotype, for example, contains considerable amounts of thymol and *γ*-terpinene (10 and 4 %, respectively) while no pure thymol chemotype was identified for this thesis. However, it should be noted that the thymol chemotype is quite rare and seldom found in natural thyme populations. Thompson et al. (2003) previously showed that the established epistatic series which controls the chemotype formation in *T. vulgaris* is not as strict as previously thought (Vernet et al. 1986). Our results are in agreement with this showing that the presence of substantial amounts of foreign monoterpenoids in a thyme plant from one chemotype is quite common. These observed deviations could arise from the more

sensitive analytical methods used, the incomplete penetrance of the epistatic series, or the genetic exchange with other chemotypes. Such variation within the major chemotypes could have a positive influence on the fitness of these plants in connection to herbivore pressure. The common garden snail *Helix aspersa*, for example, was shown to prefer the non-phenolic chemotypes, and in particular the linalool chemotype, over the phenolic chemotypes (Linhart and Thompson 1995). In this regard a plant of the linalool chemotype with substantial amounts of thymol or carvacrol could have a considerable advantage over a linalool plant which lacks these foreign monoterpenes.

Properties and phylogeny of terpene biosynthetic genes

In the course of this thesis, the sequences of six monoterpene synthases (*Tv*TPS 1 – 3 and *Tv*TPS 5 – 7), one sesquiterpene synthase (*Tv*TPS9), and two putative cytochrome P450 monooxygenases (*Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28}), were successfully isolated from *T. vulgaris* (Chapter II). Five of the six isolated monoterpene synthases were active after heterologous expression in *E. coli*: two produced linalool (*Tv*TPS2 and *Tv*TPS3), two produced the sabinene hydrates (*Tv*TPS6 and *Tv*TPS7)(Krause et al. 2013), and one produced *γ*-terpinene as their main products. Thus several key enzymes involved in the chemotype formation in *T. vulgaris* could be identified, since linalool and the sabinene hydrates define two of the seven known chemotypes, and since the two cytochrome P450s that were isolated are thought to catalyze the conversion of *γ*-terpinene to the phenolic monoterpene alcohols, thymol and carvacrol, which are defining monoterpenes of two more chemotypes in this species. (*E*)-*β*-Caryophyllene and germacrene D, two major sesquiterpenes found in the essential oil of all thyme chemotypes, are produced by two separate terpene synthases, one of which, the (*E*)-*β*-caryophyllene synthase *Tv*TPS9, could be successfully isolated and expressed, while the other could only be partially isolated (*Tv*TPS8). It should be noted that, while no sesquiterpene chemotype has been described for *T. vulgaris*, several such chemotypes are known in other *Thymus* species, for instance in *T. praecox* and *T. serpyllum* (Schmidt et al. 2004).

A phylogenetic analysis of the isolated terpene synthases (TPS), which belong to the class I terpenoid synthases based on their general structural organization, places them in the TPS-b subfamily, in close proximity to other Lamiaceae monoterpene synthase sequences and in particular sequences from members of the Menthae tribe (*Mentha*, *Salvia*, *Thymus*, and *Origanum*). Meanwhile, both sesquiterpene synthase sequences are found in close proximity

of the few other known Lamiaceae sesquiterpene synthases which cluster in the TPS-a subfamily. Previous studies have demonstrated that TPS from a certain taxon are generally more similar to TPS from phyletically close taxa than to those of phyletically more distant taxa regardless of product formed. This appears to be true for the *T. vulgaris* TPS as well. However, the thyme TPS also appear to cluster by their product specificity. For instance, the two linalool synthases *Tv*TPS2 and *Tv*TPS3 have identities on the amino acid level to other *T. vulgaris* synthases that range between 50 – 65 %. However, they have identities between 67 and 79 % to the linalool and geraniol synthases of species from the *Mentha* and *Perilla* genera. Another example is the *γ*-terpinene synthase *Tv*TPS1 which shares only 51 to 64 % and 50 to 62 % identity with the other monoterpene synthases from thyme and other Lamiaceae monoterpene synthases, respectively, but is 92 % identical to the *γ*-terpinene synthases from *Origanum vulgare*. Similar observations can be made for all other thyme mono- and sesquiterpene synthases. Both the *T. vulgaris* linalool and *γ*-terpinene synthase sequences support the common assumption that all angiosperm monoterpene synthases have arisen from a single monoterpene synthase ancestor rather than postulating multiple ancestors producing chemically similar products. However, the large sequence divergence between the Rutaceae *γ*-terpinene synthases, from *Citrus unshiu* and *C. limon*, and the Lamiaceae *γ*-terpinene synthases, from *O. vulgare* and *T. vulgaris*, and the occurrence of linalool synthases in four (TPS-b, TSP-e/f, TPS-d, and TPS-g) of the seven TPS subclasses are an indication that genes which encode enzymes displaying similar catalytic activities, are also the results of repeated evolution.

Sequence analysis of the mono- and sesquiterpene synthases from *T. vulgaris* verified the presence of the common TPS sequence motifs and elements, such as the well-known RRx_8W, $RxRx_6W$, DDxxD, and NSE/DTE motifs, and also indicated an interesting peculiarity of several Lamiaceae monoterpene synthases. The J-K loop together with the A-C loop forms the ceiling of the catalytic pocket of type I terpenoid synthases. However, this loop is dislocated to varying degrees in several thyme and other Lamiaceae TPS, due to short amino acid deletions of two to three amino acid residues, possibly resulting in easier access of water during substrate ionization. In the case of Lamiaceae TPS-b linalool synthases, this amino acid deletion is thought to be one of the main factors determining the catalytic activity of the enzyme, as the highly identical geraniol synthases from this TPS subfamily carry out similar reactions to linalool synthases but lack this particular deletion (Crowell et al. 2002, Masumoto et al. 2010). It is therefore likely that the *T. vulgaris* geraniol synthase, the key

monoterpene synthase in the geraniol chemotype but which has not yet been identified, will lack this amino acid deletion in the J-K loop but otherwise share a high identity with the two linalool synthases *Tv*TPS2 and *Tv*TPS3.

Although the leading role in terpenoid biosynthesis is certainly reserved for terpene synthases, cytochrome P450s are also known to play important parts in the formation of essential oils. One of the best studied examples are a number of limonene hydroxylases from the Lamiaceae plant family which, like the deduced thyme hydroxylase candidates *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28}, belong to the CYP71 family known for its involvement in defense-related secondary metabolism, and are known to participate in monoterpene biosynthesis (Lange and Turner 2013). Since *γ*-terpinene is presumed to be converted to thymol and carvacrol by cytochrome P450 hydroxylases similar to the known limonene-hydroxylases, and since *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} were isolated from plants containing high *γ*-terpinene, thymol or carvacrol concentrations, it is likely that both enzymes are involved in the biosynthesis of the phenolic monoterpenoids thymol and carvacrol in *T. vulgaris*.

The cytochrome P450 monooxygenase enzyme family is known for its high sequence diversity and its limited number of conserved residues and sequence motifs. It is therefore noteworthy that the Lamiaceae CYP71D subgroup, including the two sequences from *T. vulgaris*, displays an overall high sequence identity. Sequence analysis of the CYP71D members of the tribe Menthae (family Lamiaceae) results in the formation of two main clusters. One includes, apart from the two sequences isolated from thyme, enzymes known to mainly catalyze C6- and C2-hydroxylations such as the limonene-6-hydroxylases from mint. The other cluster is comprised of enzymes mainly catalyzing C3-hydroxylations. The high degree of sequence homology among the members of the Lamiaceae CYP71D, which comprises enzymes from six different plant species from four different genera, supports the hypothesis that the members of this subclass arose from a common ancestral cytochrome P450 enzyme rather than by convergence from several different ancestral enzymes with similar catalytic activities.

Even though not many sequence motifs or highly conserved residues are known in cytochrome P450s, six so-called substrate recognition sites (SRS) can often be identified by structural modeling of the enzymes. These SRS consist of comparatively short regions which form part of the catalytic site and are involved in the control of substrate access, binding, and

catalysis. Comparing the SRS of the different Lamiaceae CYP71D and taking the known catalytic activities of the analyzed enzymes into account, revealed an extremely high amino acid identity between enzymes catalyzing the hydroxylation at the same carbon atoms, and highlighted interesting residue candidates for site-directed mutagenesis.

Molecular basis of thyme monoterpene chemotypes

The regulation of terpene composition in *T. vulgaris* essential oil could take place on many different levels. For example, on the level of the genomic DNA, specific TPS genes could be absent in certain chemotypes while being present in others, or their copy number could vary greatly from one chemotype to the other. Or, the observed differences in chemotypes could be the result of 'loss of function' mutations in TPS genes so that the proteins are expressed but not active. The TPS transcript level could also be a target for regulation (Iijima et al. 2004a, Köllner et al. 2008, Crocoll 2011, Kissen et al. 2012) due to differences in transcription rate or transcript degradation, or post-transcriptional processing, such as alternative splicing. Further, the usage of rare codons in the transcribed sequence could slow protein translation and thereby influence protein degradation. The presence of an epistatic series however makes several conceivable modes of regulation unlikely to control the monoterpene polymorphism formation in *T. vulgaris*. Thus the absence of a certain monoterpene synthase from the genome of an individual chemotype or a 'loss of function' mutation of a certain monoterpene synthase inside the series are implausible regulation scenarios. Other modes of regulation, such as post-transcriptional modifications (e.g. phosphorylation), presence or absence of specific cofactors (e.g. Mg^{2+} or Mn^{2+}), kinetic differences between TPS, or the influence of TPS products themselves on TPS enzymes (e.g. inhibition of enzyme activity, inhibition of TPS transcription, or triggering of TPS enzyme degradation) are also unlikely to play pivotal parts in the chemotype formation in *T. vulgaris*. Even though the *γ*-terpinene synthase *Tv*TPS1 (Chapter II) itself produces none of the major monoterpenes defining the thyme chemotypes, it is nonetheless assumed to be a key player in the formation of thymol and carvacrol, since *γ*-terpinene is thought to be the precursor and should therefore be present in the genome of all chemotypes. The isolation of genomic DNA and subsequent sequence analysis showed that the gene encoding the synthase *Tv*TPS1 is present in the genome of all tested chemotypes and that the ORF, although differing slightly on the nucleotide level, code for an enzyme identical on the amino acid level (Chapter III). It also appeared that the gene is not subject to alternative splicing, the use of rare codons, or the presence of a 'loss of function' mutation which would lead to a premature stop codon. A

Southern blot analysis of three chemotypes spanning the whole spectrum of the epistatic series indicated that the number of γ-terpinene synthase gene copies is unlikely to play an important role in the control of chemotype formation in *T. vulgaris* (Chapter III). These results imply that the regulation of monoterpene biosynthesis in thyme is controlled at downstream steps, such as the level of transcription, since chemotype-relevant TPS genomic sequences are present in all members of the epistatic series and show a high similarity in coding and non-coding regions.

Transcriptional control of monoterpene synthases has previously been shown to be an important regulating factor for the production of certain monoterpenes, but not yet in the context of chemotype formation (Iijima et al. 2004a, Crocoll 2011). The investigation of the transcript levels of genes involved in *T. vulgaris* terpenoid biosynthesis (Chapter II) via RNA gel blotting (Chapter III) showed a strong correlation between the amount of the major essential oil components of the various chemotypes (Chapter IV) and the transcript abundance of the terpene synthases producing those components. For example, transcripts of the γ-terpinene synthase *Tv*TPS1 and the two putative cytochrome P450 enzymes *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D180-T_{28} correlated strongly with the amounts of γ-terpinene, thymol, carvacrol, and *p*-cymene found in the essential oil. Similar results could be observed for the two linalool synthases *Tv*TPS2 and *Tv*TPS3 and linalool, the two sabinene hydrate synthases *Tv*TPS6 and *Tv*TPS7 and their main products. It is therefore reasonable to assume that the control of essential oil composition in *T. vuglaris* and consequently the chemotype formation is achieved by regulation of transcript abundance.

Several mechanisms are known that influence transcript abundance and involve interaction either with the promoter region of a gene or a region close by. Specific proteins known as transcription factors, repressors, activators, silencers, or enhancers bind to the gene sequence and have a strong impact on the operation of the complex transcription machinery. These may be responsible for the differences in transcript levels of the terpene synthases that underlie chemotype differentiation in *T. vulgaris*. Other mechanisms which are associated with the gene sequence itself, rather than the promoter region, such as alternative splicing, rare codon usage, and variation in gene copy number, were shown to play an insignificant role in the regulation of thyme chemotype formation. The possibility that ploidy levels significantly influence the secondary metabolite biosynthesis in *T. vulgaris* chemotypes could also be ruled out with the assistance of flow cytometry. Using the *T. vulgaris* cultivar

'Deutscher Winter' as a reference, all chemotypes were shown to be diploid (Chapter I). It is therefore most likely that the genetic control of the *T. vulgaris* chemotype formation is directly or indirectly associated with the promoter region of the terpenoid biosynthetic genes. One possible control mechanism, which is particularly attractive in the context of the plant system used and with the information obtained so far, is epigenetic control. Epigenetic traits are stably heritable phenotypes, which are maintained by changes in the chromosome via DNA methylation, such as histone modifications, and histone variants, rather than alterations in the DNA sequence itself. Examination of the genomic sequence of the *γ*-terpinene synthase *Tv*TPS1 produced no evidence for a difference in the encoded protein sequence, but revealed small differences in the nucleotide sequences of the coding and non-coding regions (Chapter III). These changes in the nucleotide sequence of the gene could have an important influence on the chromatin architecture and consequently on the gene expression of *Tv*TPS1 itself. However, since no experiments directly addressing possible epigenetic control mechanisms were conducted, no firm conclusions can be drawn about the methylation status of the gene or its promoter or the influence of sequence changes on chromatin architecture.

Shoot regeneration from leaves of *Thymus vulgaris*

Genetic transformation of thyme would be a valuable tool for evaluating hypotheses on the molecular regulation of monoterpene chemotypes. However, transformation requires a method for regenerating plants after callus is transformed. Since no suitable regeneration protocol is available for *Thymus vulgaris*, we attempted to establish such a protocol (Chapter IV). To test the influence of plant hormones on the different stages of regeneration, including callogenesis, caulogenesis, and rooting, a series of different media was employed following closely an established protocol for another taxon of the Lamiaceae, lavandin (*Lavandula* x *intermedia*)(Dronne et al. 1999a). Three combinations of two different concentrations (4.5 µM and 9.0 µM) of the auxin indole-3-acetic acid (IAA) and the cytokinin 6-benzylaminopurine (BAP) were tested for callus induction efficiency on *T. vulgaris* leaf explants. Callus formation could be observed in all tested combinations, as well as occasionally on medium devoid of growth regulators. Similar to lavandin only slight differences were found between these three media in regards to callogenesis efficiency. However, the overall callus induction rate was far lower for thyme than what had been observed for lavandin. The average callus formation from thyme leaf explants after 14 days on callogenesis medium was 41 % on C1 medium (4.5 µM IAA and 4.5 µM BAP), 39 % on C2 medium (4.5 µM IAA and 9.0 µM BAP), and 56 % on C3 (9.0 µM IAA and 4.5 µM

BAP), while callus formation was observed on 91 – 99 % of lavandin leaf explants using similar auxin:cytokinin ratios. Apart from the auxin:cytokinin ratio, the choice of growth regulators is also very important. The synthetic auxin 2,4-dichlorophenoxyacetic acid (2,4-D), for example, is often used as a supplementary auxin to induce somatic embryogenesis by increasing the endogenous auxin levels of the explants. It has been successfully used for callus induction in different Lamiaceae, including lavandin (Dronne et al. 1999a), *Thymus* sp. (Sugisawa et al. 1988, Ozudogru et al. 2011), and different *Salvia* sp. (Kintzios et al. 1999), but was simultaneously shown to be less desirable for shoot induction, as it does not allow reduction in endogenous auxin levels which results in the inhibition of embryogenic development (Jiménez 2005). For this reason, 2,4-D was not used in our callus induction media, even though previous studies on secondary metabolite production of *T. vulgaris* callus had successfully used it both in callus induction medium and callus proliferation medium (Sugisawa et al. 1988, Tamura et al. 1993, Farag et al. 2001).

Following callus formation, caulogenesis – the generation of totipotent plant cells producing an unipolar structure whose vascular system is not independent and still connected to the parent tissue – could also be initiated in the *T. vulgaris* explants. Caulogenesis occurred on all bud induction media tested that were devoid of auxins but supplemented with 4.5 – 35.5 µM BAP, although with a very low success rate. As observed for lavandin regeneration, the callogenesis medium appeared to have an influence on subsequent shoot formation, as explants obtained from media with an IAA:BAP ratio of 1:1 or 1:2 fared best.

As previously stated plant regeneration, and in particular somatic embryogenesis, can vary from plant species to plant species, among genotypes within a species, and with the physiological state of the tissue used (Deo et al. 2010). However, the wide range of different growth regulators and their varying concentrations that were successfully used to either generate *T. vulgaris* callus (Sugisawa et al. 1988, Tamura et al. 1993, Farag et al. 2001) or micropropagate plantlets from cuttings (Furmanowa and Olszowska 1980, Lê 1989, Furmanowa and Olszowska 1992, Farag et al. 2001, Ozudogru et al. 2011) in combination with our results show that thyme is quite tolerant to these environmental factors. Nevertheless, it is necessary to define the ideal concentration and combination of growth regulators for each plant species, and in the case of the *T. vulgaris* monoterpene polymorphism, for each chemotype separately. This is particularly important for essential-oil accumulating plants such as the members of the Lamiaceae, as the growth regulators can influence terpenoid production in tissue cultures and also promote lethal browning. The

phenomenon of tissue browning and ensuing necrosis has been reported for regeneration attempts in several different plant families, including the Lamiaceae. Different factors have been suggested as the cause of browning, including the deleterious effects of oxidized phenols (Kim et al. 2007, Martin et al. 2007). The quinones resulting after phenol oxidation are highly toxic compounds to plant tissues, and are thus the most likely cause for the lethal browning that we too observed in our experiments. Although the lethal browning we observed was not as severe as reported for *Salvia fruticosa* and *S. officinalis*, where up to 100 % explants were lost (Kintzios et al. 1999), it is nonetheless a problem that will need to be addressed in further development of a *T. vulgaris* regeneration protocol.

7 Summary

Thymus vulgaris, the common garden thyme, is an aromatic plant known and valued for its essential oil since antiquity. As in most Lamiaceae species, the essential oil is a complex blend of mono- and sesquiterpenes which are responsible for its characteristic taste and odor, as well as for its prized pharmaceutical properties. In natural populations, the terpenoid composition of the oil, which is dominated by monoterpenoids and their derivatives, can vary strongly among individual plants which are otherwise morphologically identical. This variation takes the form of so-called chemotypes named after the dominant monoterpene alcohols produced in the glandular trichomes on the surface of the leaves: geraniol (G), α-terpineol (A), sabinene hydrate (thujanol, U), linalool (L), carvacrol (C), thymol (T), and 1,8-cineole (E). A genetically distinct epistatic series of five loci, with a set order of dominance (G > A > U > L > C > T), has been previously shown to be involved in controlling the monoterpenoid production of individual plants. Although the chemotype of a plant is defined by a single dominant monoterpene and its biosynthetically-related compounds, the essential oil usually contains multiple terpenoids, many of which occur only in trace amounts. In this study, some pivotal genes of terpenoid biosynthesis were isolated, characterized, and their role in monoterpenoid chemotype formation examined, to gain a better understanding of the control mechanisms underlying the chemical polymorphism in *T. vulgaris*.

The terpenoid composition of the essential oil of plants obtained from southern France was determined by GC-MS and allowed their designation as representatives of six of the seven known *T. vulgaris* chemotypes. Three of these, named G_1, A_{10}, and U_6, contained only trace amounts of monoterpenoids denominating other thyme chemotypes. The remaining plant lines, named L_{48}, C_{10}, T_2, and T_{28}, contain noteworthy amounts of monoterpenoids denominating other chemotypes. These two distinct patterns may be due to further chemotype differentiation at the level of minor compounds, or may be attributed to plants having been collected from populations with mixed chemotypes. The diploid nature of the *T. vulgaris* chemotypes was verified by flow cytometry with the known diploid cultivar 'Deutscher Winter' as reference.

A homology-based PCR strategy allowed the isolation of several *T. vulgaris* genes encoding for enzymes involved in terpenoid biosynthesis: seven terpene synthases and two

putative cytochrome P450 hydroxylases. Heterologous expression in *E. coli* resulted in five active monoterpene synthases, the *γ*-terpinene synthase *Tv*TPS1, the linalool synthases *Tv*TPS2 and *Tv*TPS3, and the sabinene hydrate synthases *Tv*TPS6 and *Tv*TPS7, as well as an active sesquiterpene synthase, *Tv*TPS9, which formed (*E*)-*β*-caryophyllene. The two cytochrome P450 gene sequences, *Tv*-CYP71D179-L_{48} and *Tv*-CYP71D-180-T_{28}, are thought to be involved in the reactions from *γ*-terpinene to thymol and carvacrol, respectively. Together these enzymes are responsible for the production of terpenoids defining four of the seven known chemotypes in *T. vulgaris* - two non-phenolic chemotypes L and U and the two phenolic chemotypes C and T. To determine the molecular basis of chemotype differences, we analyzed, taking the epistatic series into consideration, monoterpene biosynthetic genes from more than one chemotype via Southern- and RNA-blotting, and sequencing genomic clones. Based largely on work with the *γ*-terpinene synthase, which produces the precursor for the phenolic monoterpenes thymol and carvacrol, chemotype differences were shown not to result from differences in the sequences of the genes encoding the major monoterpene biosynthetic enzymes. Biosynthetic genes for producing the major monoterpenes of all chemotypes appear to be present in plant lines of all chemotypes. However, these genes were demonstrated to be differentially transcribed in different chemotypes. Transcript differences were shown not to be due to alternative splicing or differences in gene copy number, while post-transcriptional differences such as rare codon usage and 'loss of function' mutations were also not involved.

Genetic transformation of thyme could be a useful tool to prove hypotheses about the molecular regulation of chemotype formation. However neither a transformation nor a suitable regeneration protocol is known for any member of the *Thymus* genus. Here we established a regeneration protocol for *T. vulgaris* based on that for lavandin (*Lavandula* x *intermedia*) which makes it possible for the first time to regenerate *T. vulgaris* plantlets from leaf explants by inducing callus formation followed by subsequent shoot induction on media with different hormone concentrations. Although the success rate of callogenesis and especially shoot induction was low, an important step towards the development of a reliable regeneration and consequently a transformation protocol had been made.

8 Zusammenfassung

Thymus vulgaris, Echter Thymian oder auch Gartenthymian genannt, ist eine aromatische Pflanze, die seit der Antike für ihr etherisches Öl bekannt und geschätzt wird. Wie bei den meisten Arten der Lamiaceae ist das etherische Öl ein komplexes Gemisch von Mono- und Sesquiterpenen, welche sowohl für den charakteristischen Geschmack und Geruch verantwortlich sind, wie auch für viele hochgeschätzte pharmazeutische Eigenschaften. Die Terpenoidkomosition des Öls in natürlichen Populationen, welches von Monoterpenoiden und ihren Derivaten dominiert wird, variieren stark zwischen individuellen Pflanzen, die ansonsten identisch in ihrer Morphologie sind. Diese biochemische Spielart manifestiert sich in sogenannten Chemotypen, die nach dem dominanten Monoterpenalkohol benannte sind, welcher in den glandulären Trichomen auf der Blattoberfläche produziert wird: Geraniol (G), α-Terpineol (A), Sabinene hydrate (Thujanol, U), Linalool (L), Carvacrol (C), Thymol (T), und 1,8-Cineole (E). Vorrangegangene Untersuchungen haben deutlich gezeigt, dass eine epistatische Serie von fünf Loci, mit genetisch festgelegter Hierarchie (G > A > U > L > C > T), in der Kontrolle der Terpenoidproduktion der einzelnen Pflanzen involviert ist. Der Chemotyp einer Pflanze ist zwar durch ein einzelnes dominantes Monoterpen, sowie dessen biosynthetisch relevanten Verbindungen definiert, jedoch enthält das etherische Öl meistens eine Vielzahl von Terpenoiden, von denen viele nur in Spuren enthalten sind. Um ein besseres Verständnis für die Kontrollmechanismen, die dem Monoterpenpolymorphismus von *T. vulgaris* zugrunde liegen, zu erhalten, wurden in dieser Arbeit einige Schlüsselgene der Terpenebiosynthese isoliert, charakterisiert und ihre Rolle in der Bildung der Monoterpenechemotypen untersucht.

Die Terpenoidzusammensetzung des etherischen Öls der aus Südfrankreich stammenden Pflanzen, wurde mittels GC-MS bestimmt und erlaubte deren Ausweisung als Repräsentanten von sechs der sieben bekannten Chemotypen in *T. vulgaris*. Drei von ihnen, G_1, A_{10} und U_6, enthielten nur geringfügige Mengen an Monoterpenoiden, die benennend für andere Thymian Chemotypen sind. Die übrigen Pflanzenlinien, L_{48}, C_{10}, T_2 und T_{28}, hingegen, enthielten nicht unwesentliche Mengen an benennenden Monoterpenoiden anderer Chemotypen. Diese zwei klar ausgeprägten Muster könnten einerseits Folge einer weiterführenden Differenzierung der Chemotypen auf der Ebene von untergeordneten Verbindungen sein, sie könnten andererseits aber auch dem Sammeln von Pflanzen aus

Populationen mit gemischten Chemotypen zugeschrieben werden. Die diploide Natur der verwendeten Chemotypen wurde mittels Durchflusszytometrie und unter Verwendung des als diploid bekannten Kultivars 'Deutscher Winter' bestätigt.

Eine auf Homologie basierende PCR Strategie ermöglichte die Isolierung mehrerer Gene aus Thymian. Diese Gene kodieren mit sieben Terpensynthasen und zwei mutmaßliche Cytochrom P450 Hydroxylasen wichtige Enyzme der Terpenbiosynthese in *T. vulgaris*. Heterologe Expression in *E. coli* resultierte in fünf aktiven Monoterpenesynthasen, der *γ*-Terpinensynthase *Tv*TPS1, den zwei Linaloolsynthasen *Tv*TPS2 und *Tv*TPS3 und den zwei Sabinenhydratesynthasen *Tv*TPS6 und *Tv*TPS7, sowie einer aktiven Sesquiterpenesynthase, *Tv*TPS9, welche (*E*)-*β*-Caryophyllen produzierte. Es wird vermutet, dass die zwei Cytochrom P450 Gensequenzen, *Tv*-CYP71D179-L_{48} und *Tv*-CYP71D-180-T_{28}, an den Reaktionen von *γ*-Terpinen zu Thymol bzw. Carvacrol beteiligt sind. Zusammen sind diese Enzyme verantwortlich für die Biosynthese von Terpenoiden, welche vier der sieben bekannten Chemotypen in *T. vulgaris* definieren – den zwei nicht-phenolischen Chemotypen L und U, sowie den zwei phenolischen Chemotypen C und T.

Wir haben, unter Berücksichtigung der epistatischen Serie, die Gene, welche an der Monoterpenbiosynthese beteiligt sind, von mehr als einem Chemotyp mithilfe von Southernß und RNA-Blot-Methoden, sowie durch Sequenzierung genomische Klone analysiert, um die molekulare Grundlage der Unterschiede zwischen den verschiedenen Chemotypen zu ermitteln. Weitgehend auf der Arbeit mit der *γ*-Terpinensynthase basierend, die den Ausgangsstoff für die phenolischen Monoterpene Thymol und Carvacrol bildet, wurde gezeigt, dass die Unterschiede der Chemotypen nicht durch Sequenzabweichungen in wichtigen Monoterpenbiosynthese relevanten Genen verursacht wird. Des weiteren scheinen biosynthetisch relevante Gene, welche für die Produktion der Hauptmonoterpene verantwortlich sind, in den Pflanzenlinien aller Chemotypen präsent zu sein. Es konnte jedoch gezeigt werden, dass diese Gene in den verschiedenen Chemotypen unterschiedlich transkribiert wurden. Es konnte außerdem nachgewiesen werden, dass diese Unterschiede der Transkription nicht durch alternatives Spleißen oder durch Unterschiede in der Anzahl der Genkopien hervorgerufen werden. Auch posttranskriptionale Unterschiede, wie z.B. die Verwendung seltener Kodons oder 'loss of function' Mutationen, sind nicht in der Ausbildung von Chemotypen in Thymian involviert.

Die genetische Transformation von Thymian könnte ein nützliches Werkzeug sein um Hypothesen betreffend der molekularen Regulation der Chemotypausbildung zu überprüfen. Jedoch ist für keine einzige Art der Gattung *Thymus* ein geeignetes Regenerationsprotokoll oder gar ein Transformationsprotokoll bekannt. Basierend auf einem Regenerationsprotokoll für Lavandin (*Lavandula* x *intermedia*) haben wir ein solches für *T. vulgaris* etabliert. Erstmals war es möglich, Pflänzchen von *T. vulgaris* aus Blättern zu regenerieren, indem Kallusbildung mit nachfolgender Sprossinduktion auf Medien mit unterschiedlichen Hormonkonzentrationen erzeugt wurde. Auch wenn die Erfolgsrate von Kallogenesis und insbesondere der Sprossinduktion sehr niedrig war, wurde ein wichtiger Schritt gemacht, hin zur Entwicklung einer zuverlässigen Regeneration, und mitfolgend eines Transformationsprotokolls.

9 References

Aaron, J.A. and Christianson, D.W. (2010) Trinuclear metal clusters in catalysis by terpenoid synthases. *Pure and Applied Chemistry*, **82**, 1585-1597.

Adzet, T., Granger, R., Passet, J. and San Martin, R. (1977) Le polymorphisme chimique dans le genre *Thymus*: sa signification taxonomique. *Biochemical Systematics and Ecology*, **5**, 269 - 272.

Affonso, V.R., Bizzo, H.R., Lage, C.L.S. and Sato, A. (2009) Influence of growth regulators in biomass production and volatile profile of *in vitro* plantlets of *Thymus vulgaris* L. *Journal of Agricultural and Food Chemistry*, **57**, 6392-6395.

Aharoni, A., Jongsma, M., Kim, T.-Y., Ri, M.-B., Giri, A., Verstappen, F., Schwab, W. and Bouwmeester, H. (2006) Metabolic engineering of terpenoid biosynthesis in plants. *Phytochemistry Reviews*, **5**, 49-58.

Amiot, J., Salmon, Y., Collin, C. and Thompson, J.D. (2005) Differential resistance to freezing and spatial distribution in a chemically polymorphic plant *Thymus vulgaris*. *Ecology Letters*, **8**, 370-377.

Arnold, K., Bordoli, L., Kopp, J. and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195-201.

Aubourg, Aubourg, S., Lecharny, Lecharny, A., Bohlmann and Bohlmann, J. (2002) Genomic analysis of the terpenoid synthase (*At*TPS) gene family of *Arabidopsis thaliana*. *Molecular Genetics and Genomics*, **267**, 730-745.

Baek, J.-M., Han, P., Iandolino, A. and Cook, D. (2008) Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice. *Plant Molecular Biology*, **67**, 499-510.

Bakkali, F., Averbeck, S., Averbeck, D. and Idaomar, M. (2008) Biological effects of essential oils – A review. *Food and Chemical Toxicology*, **46**, 446-475.

Bennett, I.J., McDavid, D.A.J. and McComb, J.A. (2003) The influence of ammonium nitrate, pH and indole butyric acid on root induction and survival in soil of micropropagated *Eucalyptus globulus*. *Biologia Plantarum*, **47**, 355-360.

Berger, S.L., Kouzarides, T., Shiekhattar, R. and Shilatifard, A. (2009) An operational definition of epigenetics. *Genes and Development*, **23**, 781-783.

Bertea, C., Schalk, M., Mau, C.J.D., Karp, F., Wildung, M.R. and Croteau, R. (2003) Molecular evaluation of a spearmint mutant altered in the expression of limonene hydroxylases that direct essential oil monoterpene biosynthesis. *Phytochemistry*, **64**, 1203-1211.

Bittner, S. (2006) When quinones meet amino acids: chemical, physical and biological consequences. *Amino Acids*, **30**, 205-224.

Blumenthal, T. and Steward, K. (1997) RNA processing and gene structure. In *C. elegans II* (Riddle, D.L. ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 117-145.

Boeckelmann, A. (2008) Monoterpene production and regulation in lavenders (*Lavandula angustifolia* and *Lavandula* x *intermedia*). In *College of Graduate Studies - Interdisciplinary Graduate Studies*. Okanagan Campus, Kelowna, Canada: University of British Columbia.

Bohlmann, J. and Keeling, C.I. (2008) Terpenoid biomaterials. *The Plant Journal*, **54**, 656-669.

Bouhouche, N. and Ksiksi, T. (2007) An efficient in vitro plant regeneration system for the medicinal plant *Teucrium stocksianum* Boiss. *Plant Biotechnology Reports*, **1**, 179-184.

Bouwmeester, H.J., Gershenzon, J., Konings, M.C.J.M. and Croteau, R. (1998) Biosynthesis of the monoterpenes limonene and carvone in the fruit of caraway: I. Demonstration of enzyme activities and their changes with development. *Plant Physiology*, **117**, 901-912.

Britto, D.T. and Kronzucker, H.J. (2002) NH_4^+ toxicity in higher plants: a critical review. *Journal of Plant Physiology*, **159**, 567-584.

Brown, J.W.S. and Simpson, C.G. (1998) Splice site selection in plant pre-mRNA splicing. *Annual Review of Plant Physiology and Plant Molecular Biology*, **49**, 77-95.

Bruni, A. and Modenesi, P. (1983) Development, oil storage and dehiscence of peltate trichomes in *Thymus vulgaris* (Lamiaceae). *Nordic Journal of Botany*, **3**, 245-251.

Cai, Y., Jia, J.-W., Crock, J., Lin, Z.-X., Chen, X.-Y. and Croteau, R. (2002) A cDNA clone for *β*-caryophyllene synthase from *Artemisia annua*. *Phytochemistry*, **61**, 523-529.

Castillo, R., Fernández, J.-A. and Gómez-Gómez, L. (2005) Implications of carotenoid biosynthetic genes in apocarotenoid formation during the stigma development of *Crocus sativus* and its closer relatives. *Plant Physiology*, **139**, 674-689.

Chadwick, M., Trewin, H., Gawthrop, F. and Wagstaff, C. (2013) Sesquiterpenoids lactones: Benefits to plants and people. *International Journal of Molecular Sciences*, **14**, 12780 - 12805.

Chen, F., Ro, D.-K., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E. and Tholl, D. (2004) Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. *Plant Physiology*, **135**, 1956-1966.

Chen, F., Tholl, D., Bohlmann, J. and Pichersky, E. (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal*, **66**, 212-229.

Chen, X., Yauk, Y., Nieuwenhuizen, N.J., Matich, A.J., Wang, M.Y., Perez, R.L., Atkinson, R.G. and Beuning, L.L. (2010) Characterisation of an (*S*)-linalool synthase from kiwifruit (*Actinidia arguta*) that catalyses the first committed step in the production of floral lilac compounds. *Functional Plant Biology*, **37**, 232-243.

Christianson, D.W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chemical Reviews*, **106**, 3412-3442.

Coelho, N., Gonçalves, S., González-Benito, M. and Romano, A. (2012) Establishment of an in vitro propagation protocol for *Thymus lotocephalus*, a rare aromatic species of the Algarve (Portugal). *Plant Growth Regulation*, **66**, 69-74.

Conner, D.E. and Beuchat, L.R. (1984) Effects of essential oils from plants on growth of food spoilage yeasts. *Journal of Food Science*, **49**, 429-434.

Corrêa, L.d.R., Paim, D.C., Schwambach, J. and Fett-Neto, A.G. (2005) Carbohydrates as regulatory factors on the rooting of *Eucalyptus saligna* Smith and *Eucalyptus globulus* Labill. *Plant Growth Regulation*, **45**, 63-73.

Couvet, D., Bonnemaison, F. and Gouyon, P.-H. (1986) The maintenance of females among hermaphrodites: The importance of nuclear-cytoplasmic interactions. *Heredity*, **57**, 325-330.

Crocoll, C. (2011) Biosynthesis of the phenolic monoterpenes, thymol and carvacrol, by terpene synthases and cytochrome P450s in oregano and thyme. In *Biologisch-Pharmazeutische Fakultät*. Jena, Germany: Friedrich-Schiller-Universität Jena.

Crocoll, C., Asbach, J., Novak, J., Gershenzon, J. and Degenhardt, J. (2010) Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. *Plant Molecular Biology*, **73**, 587-603.

Croteau, R.B., Davis, E.M., Ringer, K.L. and Wildung, M.R. (2005) (-)-Menthol biosynthesis and molecular genetics. *Naturwissenschaften*, **92**, 562-577.

Crowell, A.L., Williams, D.C., Davis, E.M., Wildung, M.R. and Croteau, R. (2002) Molecular cloning and characterization of a new linalool synthase. *Archives of Biochemistry and Biophysics*, **405**, 112-121.

Cubas, P., Vincent, C. and Coen, E. (1999) An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*, **401**, 157 - 161.

Cullen, J., Alexander, J.C.M., Brickell, C.D., Edmondson, J.R., Green, P.S., Heywood, V.H., Jorgensen, P.-M., Jury, S.L., Knees, S.G., Maxwell, H.S., Miller, D.M., Robson, N.K.B., Walters, S.M. and Yeo, P.F. (2000) *The European Garden Flora, vol. VI Dicotyledons (Part IV): Loganiaceae to Compositae.*: Royal Horticultural Society & Cambridge University Press.

Dan, Y. (2008) Biological functions of antioxidants in plant transformation. *In Vitro Cellular Developmental Biology-Plant*, **44**, 149-161.

Daneshvar-Royandezagh, S., Khawar, K.M. and Ozcan, S. (2009) *In vitro* micropropagation of garden thyme (*Thymbra spicata* L. var. *spicata* L.) collected from southeastern turkey using cotyledon node. *Biotechnology & Biotechnological Equipment*, **23**, 1319-1321.

De Klerk, G.-J. (2002) Rooting of microcuttings: Theory and practice. *In Vitro Cellular Developmental Biology - Plant*, **38**, 415-422.

Debergh, P., Aitken-Christie, J., Cohen, D., Grout, B., Arnold, S., Zimmerman, R. and Ziv, M. (1992) Reconsideration of the term 'vitrification' as used in micropropagation. *Plant Cell, Tissue and Organ Culture*, **30**, 135-140.

Degenhardt, J., Köllner, T.G. and Gershenzon, J. (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, **70**, 1621-1637.

Delph, L.F., Touzet, P. and Bailey, M.F. (2007) Merging theory and mechanism in studies of gynodioecy. *Trends in Ecology and Evolution*, **22**, 17-24.

Demissie, Z., Cella, M., Sarker, L., Thompson, T., Rheault, M. and Mahmoud, S. (2012) Cloning, functional characterization and genomic organization of 1,8-cineole synthases from *Lavandula*. *Plant Molecular Biology*, **79**, 393-411.

Deo, P.C., Tyagi, A.P., Taylor, M., Harding, R. and Becker, D. (2010) Factors affecting somatic embryogenesis and transformation in modern plant breeding. *The South Pacific Journal of Natural and Applied Sciences*, **28**, 27-40.

Dicko, M.H., Hilhorst, R., Gruppen, H., Traore, A.S., Laane, C., van Berkel, W.J.H. and Voragen, A.G.J. (2002) Comparison of content in phenolic compounds, polyphenol oxidase, and peroxidase in grains of fifty *Sorghum* varieties from Burkina Faso. *Journal of Agricultural and Food Chemistry*, **50**, 3780-3788.

Diemer, F., Jullien, F., Faure, O., Moja, S., Colson, M., Matthys-Rochon, E. and Caissard, J.C. (1998) High efficiency transformation of peppermint (*Mentha* × *piperita* L.) with *Agrobacterium tumefaciens*. *Plant Science*, **136**, 101-108.

Dronne, S., Jullien, F., Caissard, J.-C. and Faune, O. (1999a) A simple and efficient method for *in vitro* shoot regeneration from leaves of lavandin (*Lavandula* x *intermedia* Emeric ex Loiseleur). *Plant Cell Reports*, **18**, 429-433.

Dronne, S., Moja, S., Jullien, F., Berger, F. and Caissard, J.-C. (1999b) *Agrobacterium*-mediated transformation of lavandin (*Lavendula* x *intermedia* Emeric ex Loiseleur). *Transgenic Research*, **8**, 335-347.

Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., Boland, W. and Gershenzon, J. (2005) The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 933-938.

Dudareva, N., Klempien, A., Muhlemann, J.K. and Kaplan, I. (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytologist*, **198**, 16-32.

Ehlers, B. and Thompson, J. (2004a) Do co-occurring plant species adapt to one another? The response of *Bromus erectus* to the presence of different *Thymus vulgaris* chemotypes. *Oecologia*, **141**, 511-518.

Ehlers, B.K., Grøndahl, E., Ronfort, J. and Bataillon, T. (2012) "Ménage à trois": the presence/absence of thyme shapes the mutualistic interaction between the host plant *Medicago truncatula* (Fabaceae) and its symbiotic bacterium *Sinorhizobium meliloti*. *Ecology and Evolution*, **2**, 1676-1681.

Ehlers, B.K. and Thompson, J.D. (2004b) Temporal variation in sex allocation in hermaphrodites of gynodioecious *Thymus vulgaris* L. *Journal of Ecology*, **92**, 15-23.

El Tamer, M.K., Lücker, J., Bosch, D., Verhoeven, H.A., Verstappen, F.W.A., Schwab, W., van Tunen, A.J., Voragen, A.G.J., de Maagd, R.A. and Bouwmeester, H.J. (2003) Domain swapping of *Citrus limon* monoterpene synthases: Impact on enzymatic activity and product specificity. *Archives of Biochemistry and Biophysics*, **411**, 196-203.

Elena-Rosselló, J.A. (1981) Cytotaxonomic and evolutionary studies in *Thymus* (Labiatae): Relationships of the members of section *Thymus* Jalas (1). *Anales Jardín Botánico de Madrid*, **38**, 51 - 59.

Elmore, H.W., Samples, B., Sharma, S. and Harrison, M. (1990) Influence of cultural and physiochemical factors on ascorbate stability in plant tissue culture media. *Plant Cell, Tissue and Organ Culture*, **20**, 131-135.

Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, **2**, 953-971.

Emanuelsson, O., Nielsen, H. and Heijne, G.V. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science*, **8**, 978-984.

Ennajdaoui, H., Vachon, G., Giacalone, C., Besse, I., Sallaud, C., Herzog, M. and Tissier, A. (2010) Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing *cis*-regions. *Plant Molecular Biology*, **73**, 673-685.

Erdağ, B.B. and Yürekli, A.K. (2000) In vitro propagation of *Thymus sipyleus* Boiss. (Lamiaceae). *Turkish Journal of Biology*, **24**, 81-86.

Fahn, A. (1988) Secretory tissues in vascular plants. *New Phytologist*, **108**, 229-257.

Fang, Q., Qiao, Y.-J. and Wang, H.-H. (2008) Effect of browning inhibitors on browning and the secondary metabolites in callus subculture of *Scutellaria baicalensis*. *Journal of Jilin Agricultural University*, **30**, 808-812.

Farag, R.S., Ahmed, F.A., Rady, M.R., Ghaneum, S.A. and Mater, M.A. (2001) Effect of some growth regulators and osmoregulation on the growth and essential oil production of thyme tissue culture. *Egyptian Journal of Horticulture*, **28**, 431-445.

Feyereisen, R. (2012) Insect CYP genes and P450 enzymes. In *Insect molecular biology and biochemistry* Gilbert, L.I. ed: Elsevier, pp. 236-316.

Fraternale, D., Giamperi, L., Ricci, D., Rocchi, M.B.L., Guidi, L., Epifano, F. and Marcotullio, M.C. (2003) The effect of triacontanol on micropropagation and on secretory system of *Thymus mastichina*. *Plant Cell, Tissue and Organ Culture*, **74**, 87-97.

Furmanowa, M. and Olszowska, O. (1980) *Thymus vulgaris* L. propagation through tissue culture. *Acta Poloniae Pharmaceutica*, **37**, 243-247.

Furmanowa, M. and Olszowska, O. (1992) Micropropagation of thyme (*Thymus vulgaris* L.). *Biotechnology in Agriculture and Forestry*, **19**, 230-243.

Gang, D.R. (2005) Evolution of flavors and scents. *Annual Review of Plant Biology*, **56**, 301-325.

George, E.F., Hall, M.A. and De Klerk, G.-J. (2008) Plant propagation by tissue culture. - The background. In *Plant Cell, Tissue and Organ Culture*: Springer Netherlands, pp. 353-355.

Gershenzon, J. and Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology*, **3**, 408-414.

Gershenzon, J., McCaskill, D., Rajanonarivony, J.I.M., Mihaliak, C., Karp, F. and Croteau, R. (1992) Isolation of secretory cells from plant glandular trichomes and their use in biosynthetic studies of monoterpenes and other gland products. *Analytical Biochemistry*, **200**, 130-138.

Gershenzon, J., McConkey, M. and Croteau, R. (2000) Regulation of monoterpene accumulation in leaves of peppermint. *Plant Physiology*, **122**, 205-213.

Gog, L., Berenbaum, M.R., DeLucia, E.H. and Zangerl, A.R. (2005) Autotoxic effects of essential oils on photosynthesis in parsley, parsnip, and rough lemon. *Chemoecology*, **15**, 115-119.

Gotoh, O. (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *Journal of Biological Chemistry*, **267**, 83-90.

Granger, R. and Passet, J. (1971) Types chimiques de l'espèce *Thymus vulgaris* L. *Comptes Rendus de l'Académie des Sciences*, **273**, 2350 - 2353.

Granger, R. and Passet, J. (1973) *Thymus vulgaris* spontane de France: Races chimiques et chomtaxonomie. *Phytochemistry*, **12**, 1683-1691.

Gutiérrez-Alcalá, G., Calo, L., Gros, F., Caissard, J.-C., Gotor, C. and Romero, L.C. (2005) A versatile promoter for the expression of proteins in glandular and non-glandular tichomes from a variety of plants. *Journal of Experimental Botany*, **56**, 2487-2494.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

Hamberger, B. and Bak, S. (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**.

Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A. and Deisenhofer, J. (1995) Structure and function of cytochromes P450: A comparative analysis of three crystal structures. *Structure*, **3**, 41-62.

Helander, I.M., Alakomi, H.-L., Latva-Kala, K., Mattila-Sandholm, T., Pol, I., Smid, E.J., Gorris, L.G.M. and von Wright, A. (1998) Characterization of the action of selected essential oil components on gram-negative bacteria. *Journal of Agricultural and Food Chemistry*, **46**, 3590-3595.

Henery, M.L., Moran, G.F., Wallis, I.R. and Foley, W.J. (2007) Identification of quantitative trait loci influencing foliar concentrations of terpenes and formylated phloroglucinol compounds in *Eucalyptus nitens*. *New Phytologist*, **176**, 82-95.

Hoelscher, D., Williams, D., Wildung, M. and Croteau, R. (2003) A cDNA clone for 3-carene synthase from *Salvia stenophylla*. *Phytochemistry*, **62**, 1081-1086.

Hosoi, M., Ito, M., Yagura, T., Adams, R. and Honda, G. (2004) cDNA Isolation and functional expression of myrcene synthase from *Perilla frutescens*. *Biological Pharmaceutical Bulletin*, **27**, 1979-1985.

Hull-Sanders, H.M., Johnson, R.H., Owen, H.A. and Meyer, G.A. (2009) Effects of polyploidy on secondary chemistry, physiology, and performance of native and invasive genotypes of *Solidago gigantea* (Asteraceae). *American Journal of Botany*, **96**, 762-770.

Hyatt, D.C. and Croteau, R. (2005) Mutational analysis of a monoterpene synthase reaction: Altered catalysis through directed mutagenesis of (−)-pinene synthase from *Abies grandis*. *Archives of Biochemistry and Biophysics*, **439**, 222-233.

Iijima, Y., Davidovich-Rikanati, R., Fridman, E., Gang, D.R., Bar, E., Lewinsohn, E. and Pichersky, E. (2004a) The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropens in the peltate glands of three cultivars of basil. *Plant Physiology*, **136**, 3724-3736.

Iijima, Y., Gang, D.R., Fridman, E., Lewinson, E. and Pichersky, E. (2004b) Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiology*, **134**, 370-379.

Ingvarsson, P.K. and Street, N.R. (2011) Association genetics of complex traits in plants. *New Phytologist*, **189**, 909-922.

Jiménez, V.M. (2005) Involvement of plant hormones and plant growth regulators on *in vitro* somatic embryogenesis. *Plant Growth Regulation*, **47**, 91-110.

Jonasson, P., Liljeqvist, S., Nygren, P.-A.k. and Ståhl, S. (2002) Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli. *Biotechnology and Applied Biochemistry*, **35**, 91-105.

Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, **292**, 195-202.

Jullien, F., Moja, S., Bony, A., Legrand, S., Petit, C., Benabdelkader, T., Poirot, K., Fiorucci, S., Guitton, Y., Nicolè, F., Baudino, S. and Magnard, J.-L. (2014) Isolation and functional characterization of a τ-cadinol synthase, a new sesquiterpene synthase from *Lavandula angustifolia*. *Plant Molecular Biology*, **84**, 227-241.

Kampranis, S.C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N.A., Anssour, S., Dunwell, J.M., Degenhardt, J., Makris, A.M., Goodenough, P.W. and Johnson, C.B. (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: Structural insights into the evolution of terpene synthase function. *The Plant Cell Online*, **19**, 1994-2005.

Karp, F., Mihaliak, C.A., Harris, J.L. and Croteau, R. (1990) Monoterpene biosynthesis: Specificity of the hydroxylations of (-)-limonene by enzyme preparations from peppermint (*Mentha piperita*), spearmint (*Mentha spicata*), and perilla (*Perilla frutescens*) Leaves. *Archives of Biochemistry and Biophysics*, **276**, 219-226.

Katoh, S. and Croteau, R. (1998) Individual variation in constitutive and induced monoterpene biosynthesis in grand fir. *Phytochemistry*, **47**, 577-582.

Keefover-Ring, K., Thompson, J.D. and Linhart, Y.B. (2009) Beyond six scents: defining a seventh *Thymus vulgaris* chemotype new to southern France by ethanol extraction. *Flavour and Fragrance Journal*, **24**, 117-122.

Keeling, C.I. and Bohlmann, J. (2006) Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytologist*, **170**, 657-675.

Keszei, A.A., Brubaker, C.L., Carter, R., Köllner, T., Degenhardt, J. and Foley, W.J. (2010) Functional and evolutionary relationships between terpene synthases from Australian Myrtaceae. *Phytochemistry*, **71**, 844-852.

Keszei, A.A., Brubaker, C.L. and Foley, W.J. (2008) A molecular perspective on terpene variation in Australian Myrtaceae. *Australian Journal of Botany*, **56**, 197-213.

Kim, K.-H., Lee, Y.-H., Kim, D., Park, Y.-H., Lee, J.-Y., Hwang, Y.-S. and Kim, Y.-H. (2004) Agrobacterium-mediated genetic transformation of *Perilla frutescens*. *Plant Cell Reports*, **23**, 386-390.

Kim, K.-M., Kim, M., Yun, P., Chandrasekhar, T., Lee, H.-Y. and Song, P.-S. (2007) Production of multiple shoots and plant regeneration from leaf segments of fig tree (*Ficus carica* L.). *Journal of Plant Biology*, **50**, 440-446.

Kintzios, S., Nikolaou, A. and Skoula, M. (1999) Somatic embryogenesis and *in vitro* rosmarinic acid accumulation in *Salvia officinalis* and *S. fruticosa* leaf callus cultures. *Plant Cell Reports*, **18**, 462-466.

Kissen, R., Hyldbakk, E., Wang, C.W.V., Sørmo, C.G., Rossiter, J.T. and Bones, A.M. (2012) Ecotype dependent expression and alternative splicing of epithiospecifier protein (ESP) in *Arabidopsis thaliana*. *Plant Molecular Biology*, **78**, 361-375.

Ko, W., Su, C., Chen, C. and Chao, C. (2009) Control of lethal browning of tissue culture plantlets of *Cavendish banana* cv. Formosana with ascorbic acid. *Plant Cell, Tissue and Organ Culture*, **96**, 137-141.

Kokkini, S. and Vokou, D. (1989) Carvacrol-rich plants in Greece. *Flavour and Fragrance Journal*, **4**, 1-7.

Köllner, T.G., Held, M., Lenk, C., Hiltpold, I., Turlings, T.C.J., Gershenzon, J. and Degenhardt, J. (2008) A Maize (*E*)-*β*-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *The Plant Cell Online*, **20**, 482-494.

Krasnyanski, S., May, R.A., Laskutov, A., Ball, T.M. and Sink, K.C. (1999) Transformation of the limonene synthase gene into peppermint (*Mentha piperita* L.) and preliminary studies on the essential oil profiles of single transgenic plants. *Theoretical and Applied Genetics*, **99**, 676-682.

Krause, S.T., Köllner, T.G., Asbach, J. and Degenhardt, J. (2013) Stereochemical mechanism of two sabinene hydrate synthases forming antipodal monoterpenes in thyme (*Thymus vulgaris*). *Archives of Biochemistry and Biophysics*, **529**, 112-121.

Krishna, H., Sairam, R.K., Singh, S.K., Patel, V.B., Sharma, R.R., Grover, M., Nain, L. and Sachdev, A. (2008) Mango explant browning: Effect of ontogenic age, mycorrhization and pre-treatments. *Scientia Horticulturae*, **118**, 132-138.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. *Journal of Molecular Biology*, **305**, 567-580.

Kunicka-Styczyńska, A. (2011) Activity of essential oils against food-spoiling yeast. A review. *Flavour and Fragrance Journal*, **26**, 326-328.

Landmann, C., Fink, B., Festner, M., Dregus, M., Engel, K.-H. and Schwab, W. (2007) Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). *Archives of Biochemistry and Biophysics*, **465**, 417-429.

Lane, A., Boecklemann, A., Woronuk, G., Sarker, L. and Mahmoud, S. (2010) A genomics resource for investigating regulation of essential oil production in *Lavandula angustifolia*. *Planta*, **231**, 835-845.

Lange, M.B. and Turner, G.W. (2013) Terpenoid biosynthesis in trichomes—Current status and future opportunities. *Plant Biotechnology Journal*, **11**, 2-22.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.

Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, **11**, 204-220.

Law, J.A., Vashisht, A.A., Wohlschlegel, J.A. and Jacobsen, S.E. (2011) SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA Polymerase IV. *Public Library of Science Genetics*, **7**, e1002195.

Lê, C.L. (1987) Multiplication *in vitro* de l'hysope (*Hyssopus officinalis* L.). *Revue suisse de Viticulture, Arboriculture, Horticulture*, **19**, 363-367.

Lê, C.L. (1989) Microbouturage *in vitro* du thym (*Thymus vulgaris* L.). *Revue suisse de Viticulture, Arboriculture, Horticulture*, **21**, 355-358.

Leach, G. and Whiffin, T. (1989) Ontogenetic, seasonal and diurnal variation in leaf volatile oils and leaf phenolics of *Angophora costata*. *Australian Systematic Botany*, **2**, 99 - 111.

Lee, S. and Chappell, J. (2008) Biochemical and genomic characterization of terpene synthases in *Magnolia grandiflora*. *Plant Physiology*, **147**, 1017-1033.

Li, S.-W., Xue, L., Xu, S., Feng, H. and An, L. (2009) Mediators, genes and signaling in adventitious rooting. *The Botanical Review*, **75**, 230-247.

Lima, A.S., Schimmel, J., Lukas, B., Novak, J., Barroso, J., Figueiredo, A.C., Pedro, L., Degenhardt, J. and Trindade, H. (2013) Genomic characterization, molecular cloning and expression analysis of two terpene synthases from *Thymus caespititius* (Lamiaceae). *Planta*, **238**, 191-204.

Linhart, Y.B., Keefover‐Ring, K., Mooney, K.A., Breland, B. and Thompson, J.D. (2005) A chemical polymorphism in a multitrophic setting: Thyme monoterpene composition and food web structure. *The American Naturalist*, **166**, 517-529.

Linhart, Y.B. and Thompson, J.D. (1995) Terpene-based selective herbivory by *Helix aspersa* (Mollusca) on *Thymus vulgaris* (Labiatae). *Oecologia*, **102**, 126-132.

Linhart, Y.B. and Thompson, J.D. (1999) Thyme is of the essence: Biochemical polymorphism and multi-species deterrence. *Evolutionary Ecology Research*, **1**, 151-171.

Lira-Medeiros, C.F., Parisod, C., Fernandes, R.A., Mata, C.S., Cardoso, M.A. and Ferreira, P.C.G. (2010) Epigenetic variation in mangrove plants occurring in contrasting natural environment *PLoS ONE*, **5**, e10326.

Liu, J., Xia, K.-F., Zhu, J.-C., Deng, Y.-G., Huang, X.-L., Hu, B.-L., Xu, X. and Xu, Z.-F. (2006) Nightshade proteinase inhibitor IIb gene is constitutively expressed in glandular trichomes. *Plant and Cell Physiology*, **47**, 1274-1284.

Long, M., Rosenberg, C. and Gilbert, W. (1995) Intron phase correlations and the evolution of the intron/exon structure of genes. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 12495-12499.

Lota, M.-L., de Rocca Serra, D., Tomi, F. and Casanova, J. (2001) Chemical variability of peel and leaf essential oils of 15 species of mandarins. *Biochemical Systematics and Ecology*, **29**, 77-104.

Lücker, J., Schwab, W., Franssen, M.C.R., Van Der Plas, L.H.W., Bouwmeester, H.J. and Verhoeven, H.A. (2004a) Metabolic engineering of monoterpene biosynthesis: two-step production of (+)-*trans*-isopiperitenol by tobacco. *The Plant Journal*, **39**, 135-145.

Lücker, J., Schwab, W., van Hautum, B., Blaas, J., van der Plas, L.H.W., Bouwmeester, H.J. and Verhoeven, H.A. (2004b) Increased and altered fragrance of tobacco plants after metabolic engineering using three monoterpene synthases from lemon. *Plant Physiology*, **134**, 510-519.

Lücker, J., Tamer, M., Schwab, W., Verstappen, F., van der Plas, L., Bouwmeester, H. and Verhoeven, H. (2002) Monoterpene biosynthesis in lemon (*Citrus limon*). *European Journal of Biochemistry*, **269**, 3160-3171.

Lukas, B., Samuel, R. and Novak, J. (2010) Oregano or marjoram? The enzyme *γ*-terpinene synthase affects chemotype formation in the genus *Origanum*. *Israel Journal of Plant Sciences*, **58**, 211-220.

Lupien, S., Karp, F., Wildung, M. and Croteau, R. (1999) Regiospecific cytochrome P450 limonene hydroxylases from mint (Mentha) species: cDNA isolation, characterization, and functional expression of (-)-4*S*-limonene-3-hydroxylase and (-)-4*S*-Limonene-6-hydroxylase. *Archives of Biochemistry and Biophysics*, **368**, 181-192.

Magyar-Tábori, K., Dobránszki, J., Teixeira da Silva, J., Bulley, S. and Hudák, I. (2010) The role of cytokinins in shoot organogenesis in apple. *Plant Cell, Tissue and Organ Culture*, **101**, 251-267.

Mahmoud, S.S. and Croteau, R. (2003) Menthofuran regulates essential oil biosynthesis in peppermint by controlling a downstream monoterpene reductase. *Proceedings of the National Academy of Sciences*, **100**, 14481-14486.

Mahmoud, S.S. and Croteau, R.B. (2001) Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proceedings of the National Academy of Sciences*, **98**, 8915-8920.

Mahmoud, S.S., Williams, M. and Croteau, R. (2004) Cosuppression of limonene-*3*-hydroxylase in peppermint promotes accumulation of limonene in the essential oil. *Phytochemistry*, **65**, 547-554.

Mansuy, D. and Renaud, J.-P. (1995) Heme-thiolate proteins different from cytochromes P450 catalyzing monooxygenations. In *Cytochrome P450* (Montellano, P.O. editor), Springer US, pp. 537-574.

Martin, K., Zhang, C.-L., Slater, A. and Madassery, J. (2007) Control of shoot necrosis and plant death during micro-propagation of banana and plantains (*Musa* spp.). *Plant Cell, Tissue and Organ Culture*, **88**, 51-59.

Masumoto, N., Korin, M. and Ito, M. (2010) Geraniol and linalool synthases from wild species of perilla. *Phytochemistry*, **71**, 1068-1075.

Mau, C.J.D., Karp, F., Ito, M., Honda, G. and Croteau, R.B. (2010) A candidate cDNA clone for (-)-limonene-7-hydroxylase from *Perilla frutescens*. *Phytochemistry*, **71**, 373-379.

McConkey, M.E., Gershenzon, J. and Croteau, R. (2000) Developmental regulation of monoterpene biosynthesis in the glandular trichomes of peppermint. *Plant Physiology*, **122**, 215–223.

Mendes, M., Cristina Figueiredo, A., Margarida Oliveira, M. and Trindade, H. (2013) Essential oil production in shoot cultures versus field-grown plants of *Thymus caespititius*. *Plant Cell, Tissue and Organ Culture*, **113**, 341-351.

Mendes, M.L. and Romano, A. (1999) In vitro cloning of *Thymus mastichina* L. field-grown plants. *Acta Horticulturae*, **502**, 303-306.

Misra, P. and Chaturvedi, H.C. (1984) Micropropagation of *Rosmarinus officinalis* L. *Plant Cell, Tissue and Organ Culture*, **3**, 163-168.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, **15**, 473-497.

Nagegowda, D.A. (2010) Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters*, **584**, 2965-2973.

Neera, S., Arakawa, H. and Ishimaru, K. (1992) Tannin production in *Sapium sebiferum* callus cultures. *Phytochemistry*, **31**, 4143-4149.

Nelson, D., Koymans, L., Kamataki, T., Stegeman, J., Feyereisen, R., Waxman, D., Waterman, M., Gotoh, O., Coon, M. and Estabrook, R. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, **6**, 1 - 42.

Nelson, D. and Werck-Reichhart, D. (2011) A P450-centric view of plant evolution. *The Plant Journal*, **66**, 194-211.

Nelson, D.R., Schuler, M.A., Paquette, S.M., Werck-Reichhart, D. and Bak, S. (2004) Comparative genomics of rice and *Arabidopsis*. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiology*, **135**, 756-772.

Niu, X., Li, X., Veronese, P., Bressan, R.A., Weller, S.C. and Hasegawa, P.M. (2000) Factors affecting *Agrobacterium tumefaciens*-mediated transformation of peppermint. *Plant Cell Reports*, **19**, 304-310.

Niu, X., Lin, K., Hasegawa, P.M., Bressan, R.A. and Weller, S.C. (1998) Transgenic peppermint (*Mentha* x *piperita* L.) plants obtained by cocultivation with *Agrobacterium tumefaciens*. *Plant Cell Reports*, **17**, 165-171.

O'Maille, P.E., Tsai, M.-D., Greenhagen, B.T., Chappell, J., Noel, J.P., Dan, E.R. and Joseph, P.N. (2004) Gene Library Synthesis by Structure-Based Combinatorial Protein Engineering. In *Methods in Enzymology*: Academic Press, pp. 75-91.

O'Reilly-Wapstra, J., Freeman, J., Davies, N., Vaillancourt, R., Fitzgerald, H. and Potts, B. (2011) Quantitative trait loci for foliar terpenes in a global eucalypt species. *Tree Genetics and Genomes*, **7**, 485-498.

Oldfield, E. and Lin, F.-Y. (2012) Terpene biosynthesis: Modularity rules. *Angewandte Chemie International Edition*, **51**, 1124-1137.

Osborn, T.C., Chris Pires, J., Birchler, J.A., Auger, D.L., Jeffery Chen, Z., Lee, H.-S., Comai, L., Madlung, A., Doerge, R.W., Colot, V. and Martienssen, R.A. (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics*, **19**, 141-147.

Ozudogru, E., Kaya, E., Kirdok, E. and Issever-Ozturk, S. (2011) In vitro propagation from young and mature explants of thyme (*Thymus vulgaris* and *T. longicaulis*) resulting in genetically stable shoots. *In Vitro Cellular and Developmental Biology - Plant*, **47**, 309-320.

Padovan, A., Lanfear, R., Keszei, A., Foley, W. and Kulheim, C. (2013) Correction: Differences in gene expression within a striking phenotypic mosaic Eucalyptus tree that varies in susceptibility to herbivory. *BioMed Central Plant Biology*, **13**, 57.

Page, R.D.M. (1996) Tree View: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, **12**, 357-358.

Parra, R. and Amo-Marco, J. (1996) Effect of plant growth regulators and basal media on *in vitro* shoot proliferation and rooting of *Myrtus communis* L. *Biologia Plantarum*, **38**, 161-168.

Passet, J. (1971) *Thymus vulgaris* L. - chémotaxonomie et biogénèse monoterpénique. In *Faculté de Pharmacie*. Montpellier, France: Université de Montpellier 1.

Penfold, A.R. and Willis, J.L. (1953) Physiological forms of *Eucalyptus citriodora* Hooker. *Nature*, **171**, 883-884.

Peters, R.J. and Croteau, R.B. (2003) Alternative termination chemistries utilized by monoterpene cyclases: chimeric analysis of bornyl diphosphate, 1,8-cineole, and sabinene synthases. *Archives of Biochemistry and Biophysics*, **417**, 203-211.

Petersen, M., Abdullah, Y., Benner, J., Eberle, D., Gehlen, K., Hücherig, S., Janiak, V., Kim, K.H., Sander, M., Weitzel, C. and Wolters, S. (2009) Evolution of rosmarinic acid biosynthesis. *Phytochemistry*, **70**, 1663-1679.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8**, 785-786.

Picoli, E.A.T., Paiva, E.A.S., Xavier, A., Aguiar, R.M., Carolino, S.M.B., Fári, M.G. and Otoni, W.C. (2008) Ultrastructural and biochemical aspects of normal and hyperhydric eucalypt. *International Journal of Horticultural Science*, **14**, 61-69.

Poulos, T.L. and Meharenna, Y.T. (2007) Structures of P450 proteins and their molecular phylogeny. In *The Ubiquitous Roles of Cytochrome P450 Proteins*: John Wiley & Sons, Ltd, pp. 57-96.

Poulose, A.J. and Croteau, R. (1978a) Conversion of *γ*-terpinene to *p*-cymene and thymol in *Thymus vulgaris* L. *Archives of Biochemistry and Biophysics*, **187**, 307-314.

Poulose, A.J. and Croteau, R. (1978b) *γ*-Terpinene synthetase: A key enzyme in the biosynthesis of aromatic monoterpenes. *Archives of Biochemistry and Biophysics*, **191**, 400-411.

Radulovic, N.S., Blagojevic, P.D., Stojanovic-Radic, Z.Z. and Stojanovic, N.M. (2013) Antimicrobial plant metabolites: Structural diversity and mechanism of action. *Current Medicinal Chemistry*, **20**, 932-952.

Raha, S. and Roy, S. (2003) Efficient plant regeneration in *Holarrhena antidysenterica* Wall., from shoot segment-derived callus. *In Vitro Cellular and Developmental Biology - Plant*, **39**, 151-155.

Reddy, A.S., Rogers, M.F., Richardson, D.N., Hamilton, M. and Ben-Hur, A. (2012) Deciphering the plant splicing code: Experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. *Frontiers in Plant Science*, **3**, 18.

Richards, C.L., Verhoeven, K.J.F. and Bossdorf, O. (2012) Evolutionary significance of epigenetic variation. In *Plant Genome Diversity* (Wendel, J.F., Greilhuber, J., Dolezel, J. and Leitch, I.J. eds). Vienna: Springer-Verlag, pp. 257 - 274.

Roeder, S., Hartmann, A.-M., Effmert, U. and Piechulla, B. (2007) Regulation of simultaneous synthesis of floral scent terpenoids by the 1,8-cineole synthase of *Nicotiana suaveolens*. *Plant Molecular Biology*, **65**, 107-124.

Rupasinghe, S., Baudry, J. and Schuler, M.A. (2003) Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. *Protein Engineering*, **16**, 721-731.

Saad, N.Y., Muller, C.D. and Lobstein, A. (2013) Major bioactivities and mechanism of action of essential oils and their components. *Flavour and Fragrance Journal*, **28**, 269-279.

Sáez, F., Sánchez, P. and Piqueras, A. (1994) Micropropagation of *Thymus piperella*. *Plant Cell, Tissue and Organ Culture*, **39**, 269-272.

Sakaguchi, M., Mihara, K. and Sato, R. (1987) A short amino-terminal segment of microsomal cytochrome P450 functions both as an insertion signal and as a stop-transfer sequence. *The European Molecular Biology Organization Journal*, **6**, 2425-2431.

Sallaud, C., Rontein, D., Onillon, S., Jabès, F., Duffé, P., Giacalone, C., Thoraval, S., Escoffier, C., Herbette, G., Leonhardt, N., Causse, M. and Tissier, A. (2009) A novel pathway for sesquiterpene biosynthesis from *Z*,*Z*-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. *The Plant Cell Online*, **21**, 301-317.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, **74**, 5463-5467.

Schalk, M. and Croteau, R. (2000) A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydrocylase from a C6- to a C3-hydroxylase. *Proceedings of the National Academy of Sciences*, **97**, 11948-11953.

Schmidt, A., Bischof-Deichnik, C. and Stahl-Biskup, E. (2004) Essential oil polymorphism of *Thymus praecox* subsp. *arcticus* on the British Isles. *Biochemical Systematics and Ecology*, **32**, 409-421.

Schuler, M.A. and Rupasinghe, S.G. (2011) Molecular and structural perspectives on cytochrome P450s in plants. In *Advances in Botanical Research*: Elsevier Ltd., pp. 261-307.

Seifert, A. and Pleiss, J. (2009) Identification of selectivity-determining residues in cytochrome P450 monooxygenases: A systematic analysis of the substrate recognition site 5. *Proteins: Structure, Function, and Bioinformatics*, **74**, 1028-1035.

Shahzad, A. and Siddiqui, S.A. (2000) In vitro organogenesis in *Ocimum sanctum* L. - A multipurpose herb. *Phytomorphology*, **50**, 27-35.

Shangguan, X.-X., Xu, B., Yu, Z.-X., Wang, L.-J. and Chen, X.-Y. (2008) Promoter of a cotton fibre MYB gene functional in trichomes of *Arabidopsis* and glandular trichomes of tobacco. *Journal of Experimental Botany*, **59**, 3533-3542.

Shetty, K., Carpenter, T.L., Kwok, D., Curtis, O.F. and Potter, T.L. (1996) Selection of high phenolics-containing clones of thyme (*Thymus vulgaris* L.) using *Pseudomonas* sp. *Journal of Agricultural and Food Chemistry*, **44**, 3408-3411.

Shimada, T., Endo, T., Fujii, H., Hara, M., Ueda, T., Kita, M. and Omura, M. (2004) Molecular cloning and functional characterization of four monoterpene synthase genes from *Citrus unishiu* Marc. *Plant Science*, **166**, 49-58.

Simeon de Bouchberg, M., Allegrini, J., Bessiere, C., Attisso, M., Passet, J. and Granger, R. (1976) Propriétès microbiologiques des huiles essentielles de chimotypes de *Thymus vulgaris* Linnaeus. *Rivista Italiana Essenze Profumi Piante Officinati Aromi Saponi Cosmetici*, **58**, 527 - 236.

Simkin, A.J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., Papon, N., Meyer, S., Godet, S., St-Pierre, B., Giglioli-Guivarc'h, N., Fischer, M.J.C., Memelink, J. and Clastre, M. (2013) Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid pathway in internal phloem associated parenchyma. *Phytochemistry*, **85**, 36-43.

Simmons, D. and Parsons, R.F. (1987) Seasonal variation in the volatile leaf oils of two *Eucalyptus* species. *Biochemical Systematics and Ecology*, **15**, 209-215.

Sirim, D., Widmann, M., Wagner, F. and Pleiss, J. (2010) Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. *BMC Structural Biology*, **10**, 34.

Skała, E. and Wysokińska, H. (2004) In vitro regeneration of *Salvia nemorosa* L. from shoot tips and leaf explants. *In Vitro Cellular and Developmental Biology - Plant*, **40**, 596-602.

Small, I., Peeters, N., Legeai, F. and Lurin, C. (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics*, **4**, 1581-1590.

Stahl-Biskup, E. and Sáez, F. (2002) *Thyme - The genus Thymus* 1 edn. London and New York: Taylor & Francis.

Starks, C.M., Back, K., Champpell, J. and Noel, J.P. (1997) Structural basis for cyclic terpene biosynthesis by tabacco 5-epi-aristolochene synthase. *Science*, **277**, 1815-1820.

Steele, C.L., Katoh, S., Bohlmann, J. and Croteau, R. (1998) Regulation of oleoresinosis in grand fir (*Abies grandis*): Differential transcriptional control of monoterpene, sesquiterpene, and diterpene synthase genes in response to wounding. *Plant Physiology*, **116**, 1497-1504.

Stefanello, M.É.A., Pascoal, A.C.R.F. and Salvador, M.J. (2011) Essential oils from neotropical *Myrtaceae*: Chemical diversity and biological properties. *Chemistry and Biodiversity*, **8**, 73-94.

Sugisawa, H., Miwa, K., Matsuo, T. and Tamura, H. eds (1988) Volatile compounds produced from the cultured cells of thyme (*Thymus vulgaris* L.) Berlin: We de Gruyter & Co.

Tabiyeh, D.T., Bernard, F. and Shacker, H. (2006) Investigation of glutathione, salicylic acid and GA_3 effects on browning in *Pistacia vera* shoot tips culture. *Acta Horticulturae*, **726**, 201-203.

Takahashi, S., Yeo, Y.-S., Zhao, Y., O'Maille, P.E., Greenhagen, B.T., Noel, J.P., Coates, R.M. and Chappell, J. (2007) Functional characterization of premnaspirodiene oxygenase, a cytochrome P450 catalyzing regio- and stereo-specific hydroxylations of diverse sesquiterpene substrates. *Journal of Biological Chemistry*, **282**, 31744-31754.

Tamura, H., Takebayashi, T. and Sugisawa, H. (1993) *Thymus vulgaris* L. (thyme): *In vitro* culture and the production of secondary metabolites. *Biotechnology in Agriculture and Forestry*, **21**, 413-426.

Tanaka, N., Shimomura, K. and Ishimaru, K. (1995) Tannin production in callus cultures of *Quercus acutissima*. *Phytochemistry*, **40**, 1151-1154.

Tarayre, M., Thompson, J.D., Escarré, J. and Linhart, Y.B. (1997) Population genetic structure of the gynodioecious *Thymus vulgaris* L. (Labiatae) in southern France. *Journal of Evolutionary Biology*, **10**, 157-174.

Tawfik, A.A. and Mohamed, M.F. (2007) Regeneration of salvia (*Salvia officinalis* L.) via induction of meristematic callus. *In Vitro Cellular and Developmental Biology - Plant*, **43**, 21-27.

Theis, N. and Lerdau, M. (2003) The evolution of function in plant secondary metabolites. *International Journal of Plant Sciences*, **164**, 93-102.

Tholl, D., Chen, F., Petri, J., Gershenzon, J. and Pichersky, E. (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. *The Plant Journal*, **42**, 757-771.

Thomas, C., Meyer, D., Himber, C. and Steinmetz, A. (2004) Spatial expression of a sunflower SERK gene during induction of somatic embryogenesis and shoot organogenesis. *Plant Physiology and Biochemistry*, **42**, 35-42.

Thompson, J., Charpentier, A., Bouguet, G., Charmasson, F., Roset, S., Buatois, B., Vernet, P. and Gouyon, P.-H. (2013) Evolution of a genetic polymorphism with climate change in a Mediterranean landscape. *Proceedings of the National Academy of Sciences*, **110**, 2893-2897.

Thompson, J.D., Chalchat, J.-C., Michet, A., Linhart, Y.B. and Ehlers, B. (2003) Qualitative and quantitative variation in monoterpene co-occurrence and composition in the essential oil of *Thymus vulgaris* chemotypes. *Journal of Chemical Ecology*, **29**, 859-880.

Thompson, J.D., Gauthier, P., Amiot, J., Ehlers, B.K., Collin, C., Fossat, J., Barrios, V., Arnaud-Miramont, F., Keefover-Ring, K. and Linhart, Y.B. (2007) Ongoing adaptation to Mediterranean climate extremes in a chemically polymorphic plant. *Ecological Monographs*, **77**, 421-439.

Thompson, J.D., Manicacci, D. and Tarayre, M. (1998) Thirty-five years of thyme: A tale of two polymorphisms. *BioScience*, **48**, 805-815.

Thompson, J.D. and Tarayre, M. (2000) Exploring the genetic basis and proximate causes of female fertility advantage in gynodioecious *Thymus vulgaris*. *Evolution*, **54**, 1510-1520.

Tissier, A. (2012) Glandular trichomes: what comes after expressed sequence tags? *The Plant Journal*, **70**, 51-68.

Trapp, S.C. and Croteau, R.B. (2001) Genomic oranization of plant terpene synthases and molecular evolutionary implications. *Genetics*, **158**, 811-832.

Trilles, B.L., Bombarda, I., Bouraïma-Madjebi, S., Raharivelomanana, P., Bianchini, J.-P. and Gaydou, E.M. (2006) Occurrence of various chemotypes in niaouli (*Melaleuca quinquenervia* (Cav.) S. T. Blake) essential oil from New Caledonia. *Flavour and Fragrance Journal*, **21**, 677-682.

Tullio, V., Nostro, A., Mandras, N., Dugo, P., Banche, G., Cannatelli, M.A., Cuffini, A.M., Alonzo, V. and Carlone, N.A. (2007) Antifungal activity of essential oils against filamentous fungi determined by broth microdilution and vapour contact methods. *Journal of Applied Microbiology*, **102**, 1544-1550.

Turner, G.W., Gershenzon, J. and Croteau, R. (2000) Development of peltate glandular trichomes of peppermint. *Plant Physiology*, **124**, 665-679.

Turner, G.W., Gershenzon, G., and Croteau, R. (2000) Distribution of peltate glandular trichomes on developing leaves of peppermint. *Plant Physiology*, **124**, 655-663.

van Schie, C.C.N., Haring, M.A. and Schuurink, R.C. (2007) Tomato linalool synthase is induced in trichomes by jasmonic acid. *Plant Molecular Biology*, **64**, 251-263.

Vernet, P., Gouyon, P.H. and G., V. (1986) Genetic control of the oil content in *Thymus vulgaris* L.: A case of polymorphism in a biosynthetic chain. *Genetica*, **69**, 227-231.

Vokou, D., Kokkini, S. and Bessiere, J.-M. (1993) Geographic variation of Greek oregano (*Origanum vulgare* ssp. *hirtum*) essential oils. *Biochemical Systematics and Ecology*, **21**, 287-295.

Voo, S.S., Grimes, H.D. and Lange, B.M. (2012) Assessing the biosynthetic capabilities of secretory glands in *Citrus* peel. *Plant Physiology*, **159**, 81-94.

Vottero, E., Rea, V., Lastdrager, J., Honing, M., Vermeulen, N.E. and Commandeur, J.M. (2011) Role of residue 87 in substrate selectivity and regioselectivity of drug-metabolizing cytochrome P450 CYP102A1 M11. *Journal of Biological Inorganic Chemistry*, **16**, 899-912.

Vranová, E., Coman, D. and Gruissem, W. (2012) Structure and dynamics of the isoprenoid pathway network. *Molecular Plant*, **5**, 318-333.

Wang, E., Wang, R., DeParasis, J., Loughrin, J.H., Gan, S. and Wagner, G.J. (2001) Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. *Nature Biotechnology*, **19**, 371-374.

Wang, H., Han, J., Kanagarajan, S., Lundgren, A. and Brodelius, P. (2013) Trichome-specific expression of the amorpha-4,11-diene 12-hydroxylase (cyp71av1) gene, encoding a key enzyme of artemisinin biosynthesis in *Artemisia annua*, as reported by a promoter-GUS fusion. *Plant Molecular Biology*, **81**, 119-138.

Webb, H., Lanfear, R., Hamill, J., Foley, W.J. and Külheim, C. (2013) The yield of essential oils in *Melaleuca alternifolia* (Myrtaceae) is regulated through transcript abundance of genes in the MEP pathway. *PLoS ONE*, **8**.

Weitzel, C. and Simonsen, H. (2013) Cytochrome P450-enzymes involved in the biosynthesis of mono- and sesquiterpenes. *Phytochemistry Reviews*, 1-18.

Whittington, D.A., Wise, M.L., Urbansky, M., Coates, R.M., Croteau, R.B. and Christianson, D.W. (2002) Bornyl diphosphate synthase: Structure and strategy of carbocation manipulation by a terpenoid cyclase. *Proceedings of the National Academy of Sciences*, **99**, 15375-15380.

Williams, D.C., McGarvey, D.J., Katahira, E.J. and Croteau, R. (1998) Truncation of limonene synthase preprotein provides a fully active pseudomature form of this monoterpene cyclease and reveals the function of the amino-terminal arginine pair. *Biochemistry*, **37**, 12213-12220.

Wink, M. (2010) Biochemistry, physiology and ecological functions of secondary metabolites. In *Annual Plant Reviews, Volume 40: Biochemistry of Plant Secondary Metabolism*: Wiley-Blackwell, pp. 1–19.

Wise, M.L., Savage, T.J., Katahira, E. and Croteau, R. (1998) Monoterpene Synthases from Common Sage (*Saliva officinalis*). *The Journal of Biological Chemistry*, **273**, 14891-14899.

Wüst, M., Little, D.B., Schalk, M. and Croteau, R. (2001) Hydroxylation of limonene enantiomers and analogs by recombinant (−)-limonene 3- and 6-hydroxylases from mint (Mentha) species: Evidence for catalysis within sterically constrained active sites. *Archives of Biochemistry and Biophysics*, **387**, 125-136.

Xing, W.-C., Yun, Q., Zhao, M.-A. and Cao, X. (2007) Construction of the acceptor system of direct differentiation for the gene transformation of lavender (*Lavandula angustifolia* cv. Munstead). *Arid Zone Research*, **24**, 60-66.

Yamaura, T., Tanaka, S. and Tabata, M. (1992) Localization of the biosynthesis and accumulation of monoterpenoids in glandluar trichomes of thyme. *Planta Medica*, **58**, 153-158.

Yamazaki, S., Sato, K., Suhara, K., Sakaguchi, M., Mihara, K. and Omura, T. (1993) Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. *Journal of Biochemistry*, **114**, 652-657.

Yang, T., Stoopen, G., Yalpani, N., Vervoort, J., de Vos, R., Voster, A., Verstappen, F.W.A., Bouwmeester, H.J. and Jongsma, M.A. (2011) Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. *Metabolic Engineering*, **13**, 414-425.

Yuba, A., Honda, G., Koezuka, Y. and Tabata, M. (1995) Genetic analysis of essential oil variants in *Perilla furtescens*. *Biochemical Genetics*, **33**, 341-348.

Yuba, A., Yazaki, K., Tabata, M., Honda, G. and Croteau, R. (1996) cDNA cloning, characterization, and functional expression of 4*S*-(-)-limonene synthase from *Perilla frutescens*. *Archives of Biochemistry and Biophysics*, **332**, 280-287.

Zaina, S., Perez-Luque, E.L. and Lund, G. (2010) Genetics talks to epigenetics? The interplay between sequence variants and chromatin structure. *Current Genomics*, **11**, 359-367.

Zhang, S., Miao, H., Sun, R., Wang, X., Huang, S., Wehner, T.C. and Gu, X. (2013) Localization of a new gene for bitterness in cucumber. *Journal of Heredity*, **104**, 134-139.

10 Acknowledgments

The completion of my dissertation has been a protracted and arduous journey. Life doesn't stand still and wait until you have time to cope with it. Much has happened and changed in the time I have been entangled in this project, and it has certainly shaped me as a person. Many have, for plausible reasons, doubted whether I would finish my thesis and have questioned my commitment to it. I, on the other hand, while losing confidence so many times I've lost count, getting writer's block just as many times, getting sick, and dissertation woes in general, never lost hope and clung desperately to the project. I wasn't able to give up and 'run away', mainly because I felt that I had to finish the dissertation for myself, and partly because of all the invaluable support and help I was given by faculty, friends and family members.

At any rate, I have finished, but not alone, and am elated. At this point, many names come to my mind for a multitude of reasons; almost constantly these are a mix of professional and personal motives, which are hardly distinguishable. It is customary for dissertators to reserve their most exaggerated sentimentality and pathos, in short their purplest prose, for their acknowledgments. Such a time-honored tradition needs to be respected and if readers cringe at this section but accept the overall dissertation as good science, I will be delighted.

Space is limited and there is no way to avoid a concise selection from the multitude of individuals who have each contributed in their own way to the completion of this dissertation. My heartfelt thanks go to all of them alike, credited or uncredited.

The select group I'd like to give my thanks to, is headed first and foremost by my Professor, Jonathan Gershenzon, whose intellectual heft is matched only by his genuinely good nature and down-to-earth humanity. He guided me, with the patience of a saint, from an idea through to a complete study, and I am truly fortunate to have had the opportunity to have him as a mentor and work with him.

It was a delight to have Christoph Crocoll as a fellow doctoral student and as a bench-mate in the laboratory. Working on closely related plant species, similar enzyme families, and like scientific questions, led not only to a productive cooperation but also to a solid friendship and many fruitful and original discussions. It is due to his pioneering spirit that I tentatively put out my feelers into the wide field of cytochrome P450 enzymes.

I would like to thank Professor Jörg Degenhardt for introducing me to the multifaceted cosmos of terpene synthases and patiently guiding my first steps into this fascinating world, as well as taking Sandra Krause and Jette Schimmel under his wing and thus continuing the work on *T. vulgaris*. Sandra and Jette I have to specifically thank for their willingness to share valuable information about the enzymatic characterization of *Tv*TPS5, *Tv*TPS6, and *Tv*TPS7, without which the story presented in this dissertation would have been less interesting and incomplete.

This entire dissertation on *T. vulgaris* chemotypes would have been impossible had it not been for the generous support of Professor John Thompson who kindly provided me with a selection of thyme chemotypes plants as well as seemingly infinite information about this interesting herb and its polymorphisms. To the greenhouse staff I would like to express my gratitude for accommodating this troublesome and capricious herb and providing it with shelter and loving care, and in particular Jana Zitzmann, Udo Kornmesser, and Tamara Krügel. The work on the regeneration protocol would have remained a future endeavor had it not been for the support of Tamara Krügel and Ulrike Temp, who not only gave me access to different plant media and hormones that I needed, but most of all helped me with my first steps in this complex and fascinating field, while Susanne Junghans lent a helping hand at the sterile-bench and kept this project manageable.

The members of the Department of Biochemistry at the Max Planck Institute for Chemical Ecology, with their dedication to their work, made the environment here one of warm hospitality with an extraordinary cooperative and collaborative mindset, and all in all fantastic working conditions. I have to mention in particular Susanne Textor, Axel Schmidt, Michael Reichelt, Jan-Willem de Kraker, Tobias Köllner, and of course the technicans and Angela Schneider who are the good souls of the group and keep everything running smoothly. However, I would also like to thank my fellow doctoral students from the International Max Plank Research School – those who have moved on, those in the quagmire, and those just beginning – for their support, feedback and friendship. And in particular some students from our research group, who have become dear to me: Christoph Crocoll, Diana Koepke, Meike Burow, Christine Opitz, Raimund Nagel, Kerstin Büchel, Katharina Schramm and Alexander Schwarzkopf.

I also gratefully acknowledge the institutional support that I have received while working on this project. Be it the wizards of the IT-department, the library, and the

greenhouse, or the members of the administration, I only ever experienced unconditional help and support from them.

One thing everyone should have in this world, are true friends. Those who are prepared to put up with one's moods and are not discouraged by outbursts of frustration or month-long silences, those who poke and coax one to come out of one's shell and look at the world outside, those who are there when needed. Thank you Wil, Romain, Ute, Evi, Annemarie, Sarah and Verena for being such friends.

Finally, I'd be remiss if I didn't acknowledge the innumerable sacrifices made my insanely large, or dare I say largely insane family. Blood related or not, they supported, encouraged and helped me countless times over the years and kept my hope for post-dissertation normality alive. My greatest debt of appreciation goes to my mother, who took the blows and made innumerable sacrifices. Thank you for giving me a chance to thrive.

11 Supplementary Material

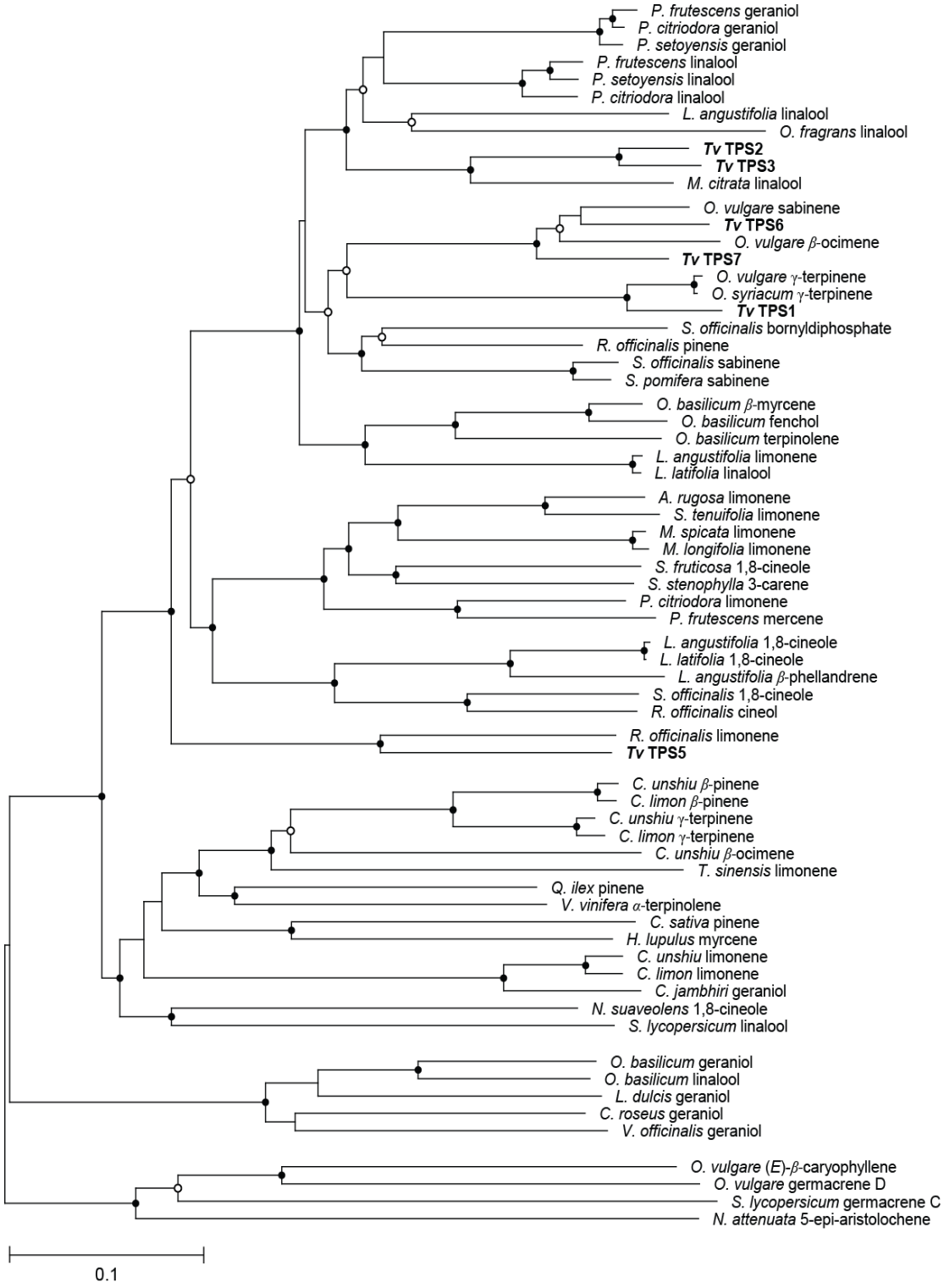


Figure S1: Dendrogram analysis showing the positioning of monoterpene synthases isolated from *T. vulgaris* and functionally related terpene synthases in a phylogenetic context. The dendrogram was constructed using the Neighbor-Joining method and rooted with lamiid sesquiterpene synthases. Bootstrap values of ≥ 80 % are indicated by black dots on the nodes, while values of 70 - 79 % are indicated by circles on the nodes (10.000 replicates). Sequences determined in this thesis are in boldface. The name of the major enzyme product is given after the abbreviation of the species. GenBank accession numbers are in parentheses. *A. rugosa – Agastache rugosa*: limonene synthase (AAL17636); *C. limon – Citrus limon*: limonene synthase (AF514289), *β*-pinene synthase (AF514288), *γ*-terpinene synthase (AAM53943); *C. sativa – Cannabis sativa*: *α*-pinene synthase (DQ839405); *C. jambhiri – Citrus jambhiri*: geraniol synthase (BAM29049); *C. roseus – Catharanthus roseus*: geraniol synthase (AFD64744); *C. unshiu – Citrus unshiu*: limonene synthase (BAD27257), *β*-ocimene synthase (BAD91046), *β*-pinene synthase (BAD27260), *γ*-terpinene synthase (BAD27259); *H. lupulus – Humulus lupulus*: myrcene synthase (EU760349); *L. angustifolia – Lavandula angustifolia*: 1,8-cineole synthase (JN701461), limonene synthase (DQ263740), linalool synthase (ABB73045), *β*-phellandrene synthase (HQ404305); *L. latifolia – Lavandula latifolia*: 1,8-cineole synthase (JN701460), linalool synthase (DQ421801); *L. dulcis – Lippia dulcis*: geraniol synthase (GU136162); *M. citrata – Mentha citrata*: linalool synthase (AAL99381); *M. longifolia – Mentha longifolia*: limonene synthase (AAD50304); *M. spicata – Mentha spicata*: limonene synthase (AAC37366); *N. attenuata – Nicotiana attenuata*: 5-epi-aristolochene synthase (AF542544); *N. suaveolens – Nicotiana suaveolens*: 1,8-cineole synthase (ABP88782); *O. basilicum – Ocimum basilicum*: fenchol synthase (AY693648), geraniol synthase (AY362553), linalool synthase (AY693647), *β*-myrcene synthase (AY693649), terpinolene synthase (AY693650); *O. fragrans – Osmanthus fragrans*: linalool synthase (ACM92062); *O. syriacum – Origanum syriacum*: *γ*-terpinene synthase (JN542829); *O. vulgare – Origanum vulgare*: (*E*)-*β*-caryophyllene synthase (GU385970), germacrene D synthase (GU385976), *γ*-terpinene synthase (GU385978), sabinene synthase (GU385980), *β*-ocimene synthase (GU385967); *P. citriodora – Perilla citriodora*: geraniol synthase (DQ088667), limonene synthase (AF233894), linalool synthase (AY917193); *P. frutescens – Perilla frutescens*: linalool synthase (AAL38029), geraniol synthase (FJ644547), mercene synthase (AF271259); *P. setoyensis – Perilla setoyensis*: geraniol synthase (FJ644545), linalool synthase (FJ644544); *Q. ilex – Quercus ilex*: pinene synthase (CAK55186); *R. officinalis – Rosmarinus officinalis*: cineol synthase (DQ839411), limonene synthase (ABD77416), pinene synthase (ABP01684); *S. fruticosa – Salvia fruticosa*: 1,8-cineole synthase (ABH07677); *S. officinalis – Salvia officinalis*: bornyldiphosphate synthase (AF051900), 1,8-cineole synthase (AF051899), sabinene synthase (AF051901); *S. pomifera – Salvia pomifera*: sabinene synthase (DQ785794); *S. stenophylla – Salvia stenophylla*: 3-carene synthase (AAM89254); *S. lycopersicum* - *Solanum lycopersicum:* germacrene C synthase (AAC39432), linalool synthase (AEM05855); *S. tenuifolia – Schizonepeta tenuifolia*: limonene synthase (AAG01140); *T. sinensis – Toona sinensis*: limonene synthase

(AB303572); *Tv*TPS – *Thymus vulgaris*: γ-terpinene synthase (*Tv*TPS1), linalool synthases (*Tv*TPS2 and *Tv*TPS3), inactive protein (*Tv*TPS5), sabinene hydrate synthases (*Tv*TPS6 and *Tv*TPS7); *V. vinifera* – *Vitis vinifera*: α-terpinolene synthase (AY572987); *V. officinalis* – *Valeriana officinalis*: geraniol synthase (KF951406).

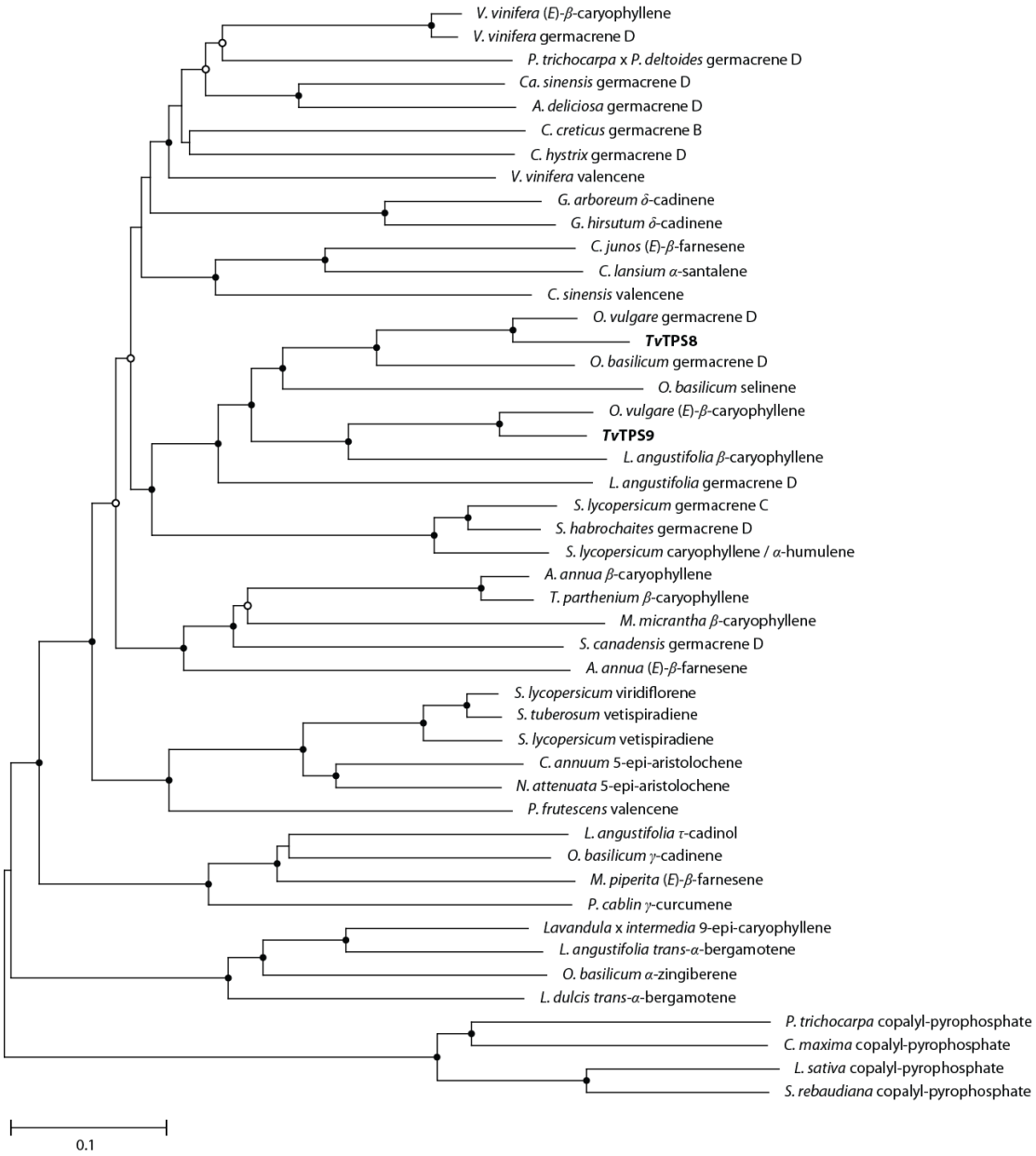


Figure S2: Dendrogram analysis showing the positioning of sesquiterpene synthases isolated from *T. vulgaris* and functionally related terpene synthases in a phylogenetic context. The dendrogram was constructed using the Neighbor-Joining method and rooted with copalyl-

pyrophosphate synthases from the TPS-c subfamily. Bootstrap values of ≥ 80 % are indicated by black dots on the nodes, while values of ≥ 60 % are indicated by circles on the nodes (10.000 replicates). Sequences determined in this thesis are in boldface. The name of the major enzyme product is given after the abbreviation of the species. GenBank accession numbers are in parentheses. *A. annua* – *Artemisia annua*: (*E*)-*β*-caryophyllene synthase (AAL79181), (*E*)-*β*-farnesene synthase (AAX39387); *A. deliciosa* – *Actinidia deliciosa*: germacrene D (AAX16121); *Ca. sinensis* – *Camellia sinensis*: putative germacrene D synthase (AFE56211); *C. annuum* – *Capsicum annuum*: 5-epi-aristolochene synthase (CAA06614); *C. creticus* – *Cistus creticus*: germacrene B synthase (ACF94469); *C. lansium* – *Clausena lansium*: *α*-santalene synthase (ADR71055); *C. junos* – *Citrus junos*: (*E*)-*β*-farnesene synthase (AAK54279); *C. hystrix* – *Citrus hystrix*: germacrene D synthase (ADX01384); *C. sinensis* – *Citrus sinensis*: valencene synthase (AAQ04608); *C. maxima* – *Cucurbita maxima*: copalyl-pyrophosphate synthase (AAD04292); *G. arboreum* – *Gossypium arboreum*: *δ*-cadinene synthase (CAA65289); *G. hirsutum* – *Gossypium hirsutum*: *δ*-cadinene synthase (AAF74977); *L. angustifolia* – *Lavandula angustifolia*: *trans*-*α*-bergamotene synthase (ABB73046), *τ*-cadinol synthase (AGL98418), (*E*)-*β*-caryophyllene synthase (AGL98419), germacrene D synthase (AGL98420); *L.* x *intermedia* – *Lavandula* x *intermedia*: 9-epi-caryophyllene synthase (AGU13712); *L. dulcis* – *Lippia dulcis*: *trans*-*α*-bergamotene synthase (AFR23371); *L. sativa* – *Lactuca sativa*: copalyl-pyrophosphate synthase (BAB12440); *M. micrantha* – *Mikania micrantha*: *β*-caroyphyllene synthase (ACN67535); *M. piperita* – *Mentha piperita*: (*E*)-*β*-farnesene synthase (AAB95209); *N. attenuata* – *Nicotiana attenuata*: 5-epi-aristolochene synthase (AAO85555); *O. basilicum* – *Ocimum basilicum*: *γ*-cadinene synthase (AAV63787), germacrene D synthase (AAV63786), selinene synthase (AAV63785), *α*-zingiberene synthase (AAV63788); *O. vulgare* – *Origanum vulgare*: (*E*)-*β*-caryophyllene synthase (ADK73616), germacrene D synthase (ADK73619); *P. cablin* – *Pogostemon cablin*: *γ*-curcumene synthase (AAS86319); *P. frutescens* – *Perilla frutescens*: valencene synthase (AAX16077); *P. trichocarpa* x *P. deltoides* – *Populus trichocarpa* x *P. deltoides*: germacrene D synthase (AAR99061); *P. trichocarpa* – *Populus trichocarpa*: copalyl-pyrophosphate synthase (EEE81383); *S. habrochaites* – *Solanum habrochaites*: germacrene D synthase (AAG41892); *S. canadensis* – *Solidago canadensis*: germacrene D synthase (AAR31145); *S. lycopersicum* – *Solanum lycopersicum*: caryophyllene/*α*-humulene synthase (NP001234766), germacrene C synthase (NP001234060), vetispiradiene synthase (AAG09949), viridiflorene synthase (AEP82773); *S. tuberosum* – *Solanum tuberosum*: vetispiradiene synthase (AAD02223); *S. rebaudiana* – *Stevia rebaudiana*: copalyl-pyrophosphate synthase (AAB87091); *T. parthenium* – *Tanacetum parthenium*: (*E*)-*β*-caryophyllene synthase (AEH41845); TvTPS – Thymus vulgaris: (*E*)-*β*-caryophyllene synthase (TvTPS9), partial germacrene D synthase (TvTPS8); *V. vinifera* – *Vitis vinifera*: (*E*)-*β*-caryophyllene (AEP17005), germacrene D (NP001268213), valencene (AAS66358).

Table S1: Oligonucleotide sequences used for gene isolation, gene expression, and synthesis of RNA- and DNA-hybridization probes.

Oligonucleotide sequences for gene expression, start and stop codons are underlined	
attB1-*Tv*TPS1-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGGCTTCCCTTAGCATGCAAGTCTC-3'
attB1-*Tv*TPS1cut-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGAGGCGCGTCTCTACCACTCGT-3'
attB2-*Tv*TPS1stop-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTCACACGTACGGCTCGAAGATG-3'
attB1-*Tv*TPS2-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGTCTGCAACCATTAGCGTA-3'
attB1-*Tv*TPS2cut-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGAGGCGTTCCGGGAACTACC-3'
attB2-*Tv*TPS2stop-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGGATATGGGTGGAACAGG-3'
attB1-*Tv*TPS3-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTGATGATCTTGTCCTCTGCAAA-3'
attB1-*Tv*TPS3cut-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTGATGCGGCGTTCCGGAAAATACC-3'
attB2-*Tv*TPS3stop-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGGATACGGCTGGAAC-3'
attB1-*Tv*TPS9-fwd	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGAATTTCCGGCATCGGTTGCT-3'
attB2-*Tv*TPS9-rev	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATTATACGGGATCAACGAGTATGGATTT-3'
Oligonucleotide sequences for DNA hybridization analysis	
*Tv*TPS1-southern-fwd	5'-TAA GAA CGA GGA CGG TAG TGA TTT-3'
*Tv*TPS1-southern-rev	5'-ATT TTC CCG AGC TCG TAG ATG ATT-3'
Oligonucleotide sequences for *Tv*TPS1 isolation from gDNA	
gDNA-*Tv*TPS1-fwd	5'-TAA GAA TTC GCT TCC CTT AGC ATG CAA GTC TCC-3'
gDNA-*Tv*TPS1-rev	5'-CAA AAG CTT TCA CAC GTA CGG CTC GAA GAT GAG GC-3'

Oligonucleotide sequences for RNA hybridization analysis	
*Tv*TPS1-fwd	5'-TCA CCG ACG CGA TTC GAA AAT GGG ACT T-3'
*Tv*TPS1-rev	5'-GGG GTC CGC CAT CGC CGT GTTC-3'
*Tv*TPS2-fwd	5'-CGC CTC TCC GGG AAA CTC GTA-3'
*Tv*TPS2-rev	5'-ACC ACG TCA CCA TTA AAT TTT GAC AGA A-3'
*Tv*TPS7-fwd	5'-CTT TTG GGC AGT TGG GCT ATT TGA GGG-3'
*Tv*TPS7-rev	5'-TCG TAA GCC AGA TCG GAG ACA AAG TT-3'
*Tv*TPS9-fwd	5'- CAC ACT ATT CAC TAG CGT TGT TCG AAG GTG GGA CAT -3'
*Tv*TPS9-rev	5'- GGT CGT AAG GAT TGG AAC AGA GGC TGG TCG T -3'
*Tv*P450-fwd	5'-GGC TTC ACC CTC CTT TCC CGA TTA TAC CAA GAC-3'
*Tv*P450-rev	5'-GGC CTG GCT TCT CCG TCA TGT CAA TGT C-3'

Table S2: Plasmid vectors utilized in this thesis.

Vector	Features	Application
pCR4®-TOPO® (Invitrogen)	bacterial cloning vector, 3957 bp, Kan^r, Amp^r, *lac*Zα, *ccd*B	Cloning of PCR products with 3' adenine-overhang
pDNR-LIB (BD Bioscience Clontech)	bacterial cloning vector, 4200 bp, Cm^r, *Sac*B	Establishing of Creator™ SMART™ cDNA library
pH9GW (O'Maille et al. 2004)	bacterial expression vector with C-terminal His(9)-tag, pET-T7 (28a)-derivative, 6990 bp, Kan^r, Cm^r, *lac*Zα, *ccd*B, T7-promoter	Overexpression of N-terminal 9x His-tag
pDONR™207 (Invitrogen)	bacterial cloning vector, 5585 bp, $Gent^r$, Cm^r, *ccd*B	Cloning of *att*B-PCR products to create Gateway® donor vector for subcloning

Table S3: Bacterial strains utilized in this thesis.

Strains	Genotype	Application
One Shot® TOP 10 (Invitrogen)	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Used for chemical transformation of DNA fragments ligated into pCR4-TOPO-TA.
One Shot® BL21(DE3) (Invitrogen)	F⁻ *omp*T *gal dcm lon hsd*S_B ($r_B^-m_B^-$) (DE3) (λ prophage with T7-RNA-polymerase gene)	Used for expression of subcloned terpene synthase genes under the control of a T7-promoter.
ElectroMAX DH5α (Invitrogen)	F⁻ Φ80*lac*ZΔM15 (*lac*ZYA-*arg*F) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_k-, m_k+) *gal*⁻ *pho*A *sup*E44 λ⁻ *thi*⁻1 *gyr*A96 *rel*A1	Used for Creator™ SMART™ cDNA library construction with the pDNR-LIB vector.

12 Eigenständigkeitserklärung

Hiermit erkläre ich, entsprechend § 5 Absatz 4 der Promotionsordnung der Biologisch-Pharmazeutischen Fakultät der Friedrich Schiller Universität Jena, das mir die geltende Promotionsordnung bekannt ist. Die vorliegende Dissertation habe ich eigenständig und nur unter Verwendung angegebener Quellen und Hilfsmittel angefertigt, wobei von Dritten übernommene Textabschnitte entsprechend gekennzeichnet wurden. Alle Personen, die einen entscheidenden Beitrag zu den Kapiteln geleistet haben, sind dort aufgeführt beziehungsweise in der Danksagung erwähnt. Die Hilfe eines Promotionsberaters wurde nicht in Anspruch genommen noch haben Dritte geldwerte Leistungen für Arbeiten im Zusammenhang mit der vorliegenden Disseration erhalten. Zu keinem früheren Zeitpunkt wurde diese Disseration, eine in wesentlichen Teilen ähnliche Arbeit oder eine andere Abhandlung bei einer Hochschule als Disseration eingereicht.

Julia Asbach

Jena, den

13 Curriculum Vitae

Personal

Name: Julia Nina Christina Asbach
Date of birth: 03rd May 1977
Place of birth: Hamburg, Germany

Education

2003 – 2014	PhD at the Friedrich-Schiller-University of Jena and the Max Planck Institute for Chemical Ecology, Jena, Germany Department of Biochemistry Supervisor: Prof. Dr. Jonathan Gershenzon
2002 – 2003	Diploma at the Ernst-Moritz-Arndt-University of Greifswald, Germany Institute of Microbiology "Alanin-Aminotransferase in *Saccharomyces cerevisiae*" Supervisor: Prof. Dr. Rüdiger Bode
1997 – 2003	Studies in Biology at the Ernst-Moritz-Arndt-University of Greifswald, Germany Majors: Biochemistry, Plant physiology, Genetics
1997	Abitur at the Sankt-Ansgar-Schule, Hamburg, Germany

Publications

Krause, S. T., Köllner, T., Asbach, J., Degenhardt, J. (2013). Stereochemical mechanism of two sabinene hydrate synthases forming antipodal monoterpenes in thyme (*Thymus vulgaris*). *Archives of Biochemistry and Biophysics* **529**(2):112-121.

Crocoll, C., Asbach, J., Novak, J., Gershenzon, J., Degenhardt, J. (2010). Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. *Plant Molecular Biology* **73:**587-603.

Poster Presentations

Asbach, J., Thompson, J., Gershenzon, J., Degenhardt, J. (2007) The role of terpene synthase expression in chemotype formation in thyme. *IMPRS Evaluation Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Asbach, J., Thompson, J., Gershenzon, J., Degenhardt, J. (2007) The role of terpene synthase expression in chemotype formation in thyme. *TERPNET 2007 – 8th international meeting on biosynthesis and function of isoprenoids in plants, microorganisms and parasites*, Strasbourg, France

Asbach, J., Thompson, J., Gershenzon, J., Degenhardt, J. (2007) The role of terpene synthase expression in the formation of chemotypes in thyme. *6th IMPRS Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Asbach, J., Thompson, J., Gershenzon, J., Degenhardt, J. (2006) Role of terpene synthase expression in the formation of chemotypes in thyme. *5th IMPRS Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Asbach, J., Crocoll, C., Thompson, J., Gershenzon, J., Degenhardt, J. (2006) Elucidating the biochemical basis of monoterpene polymorphism in *Thymus vulgaris* L. *ICOB-5 and ISCNP-25 IUPAC: International Conference on Biodiversity and Natural Products*, Kyoto, Japan

Asbach, J., Gershenzon, J., Degenhardt, J. (2006) Chemical polymorphism and terpene biosynthesis in *Thymus vulgaris* L. *ICE Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Asbach, J., Gershenzon, J., Degenhardt, J. (2005) The biochemical and genetical regulation of the polychemism in *Thymus vulgaris*. *3rd IMPRS Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Asbach, J., Gershenzon, J., Degenhardt, J. (2005) Monoterpene biosynthesis in the mint family. *ICE Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Oral Presentations

Asbach, J. (2008) Terpene biosynthesis in *Thymus vulgaris* L. *7th IMPRS Symposium* / *Max Planck Institute for Chemical Ecology*, Dornburg, Germany

Asbach, J. (2006) Monoterpene biosynthesis in common garden thyme (*Thymus vulgaris*) *4th IMPRS Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Asbach, J. (2005) Chemical polymorphism and terpene biosynthesis in *Thymus vulgaris* L. *5th Kurt-Mothes-Doktoranden-Workshop Sekundärstoffwechsel* / *Leibniz Institute for Plant Biochemistry (IPB)*, Halle (Saale), Germany

Asbach, J. (2005) The chemical polymorphism in *Thymus vulgaris* L. *2nd IMPRS Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany

Asbach, J. (2004) Polychemism in *Thymus vulgaris* L. *1st IMPRS Symposium* / *Max Planck Institute for Chemical Ecology*, Jena, Germany